Queen Mary University of London
School of Economics and Finance

A Financial Analysis of New Econometric Techniques

A Thesis submitted for the degree of
Doctor Philosophy (Ph.D.) in Finance

Author:
Fabio Calonaci

Supervisor:
Prof. Liudas Giraitis

September 2019

Declaration of Authorship

I wish to declare:

No part of this doctoral dissertation, titled "A Financial Analysis of New Econometric Techniques "and submitted to Queen Mary University of London in pursuance of the degree of Doctor of Philosophy (Ph.D) in Finance, has been presented to any other University.
Parts of Chapter 2 were undertaken as joint work with Prof. Richard Baillie and Prof. George Kapetanios.
Parts of Chapter 3 were undertaken as joint work with Prof. Richard Baillie, Dr. Dooyeon Cho and Dr. Seunghwa Rho.

Abstract

The predictability of asset returns is one of the most controversial topic in financial literature other than critical issue for portfolio managers. Over the last decades, the topic has been investigated from two different points of view: one more empirical with new models and one more theoretical with new estimation techniques. Understanding the complexity of the topic, we decided to contribute to both of these approaches. On one side we propose a new model of financial volatility, while on the other we develop a new econometric methodology for better capturing the risk of portfolios, which are both key elements in the analysis of returns.

The first part of the Thesis suggests a compressive approach for capturing the volatility of assets returns. It extends the Heterogeneous Autocorrelation model, Corsi (2009), HAR, decomposing the volatility into its principal features: slow decay of the autocorrelation function, asymmetric behaviour with returns and all the facets of volatility jumps. The empirical forecasting exercise shows remarkable improvements in the precision of the forecasts.

The second part of the Thesis introduces a new hierarchical methodology for estimating dynamic pricing models and increasing their performances. The method, based on the Fama and MacBeth (1973) approach and developed in the classical kernel regression framework, employs a flexible algorithm for the selection of the bandwidth. Wide empirical evidence is provided in support of the methodology and its ability to produce a more accurate description of the systematic risk. The last part of the Thesis sheds light on the relationship among different long memory models. The chapter investigates the role of long memory parameters in addition to the classical HAR approach finding that its importance varies across the assets. The HAR restricted ARFIMA model seems to be a good approximation of the dynamic structure of several Realized Volatility series.

Acknowledgements

This work is the final outcome of a journey started almost 6 years ago, in which I have challenged myself professionally and personally. Along the journey, I have often questioned whether I would even make it to the end. Well, I wouldn't gotten so far without the help, support and endurance of many people.

Firstly, I thank my supervisor Prof. Liudas Giraitis for all the time he has dedicated to me and for guiding me throughout my Ph.D. I have deeply benefited from his inputs. Secondly, I thank my co-author Prof. Richard Baillie. He has been positive influence who had a huge impact on my research interests guided me with enthusiasm providing me precious advices. I thank Prof. George Kapetanios, who has been incredibly helpful for all my projects and constant support in all my non academic decisions.

I also would like to thank Lollo and Eli who played a huge role in the last 6 years; in the rough times to help and encourage me and in the good ones, to laugh and share a beer. I'm sure I wouldn't be an inch of what I'm now without you. Elisa for being the perfect flatmate and becoming my London family; I will miss you. Alessio, Daniele, Edoardo and Marco to make going to the office easy and pleasant like going to an Italian bar and to all the QM Ph.D students you all made my time in London great an I will always be grateful.

Thank you Sergy; despite the distance, you have always been there. Thank you Fagio, we have been friends for almost three decades and you help me not to forget my roots. To many others who played an important role throughout this journey, and the little thing I can do is to name them: Claudio, Prof. Gallo, Luca, Marco, Silvia, Valerio.

Finally, a special thank goes to my family, Babbo, Mamma, Gio and Ire; although it was hard to see me via FaceTime, you always have been supporting with your unconditional love, I would not make it without your love. I love you even if I do not show it often. Mimma for being always there and taking care of my Babbo. A big hug goes also to my Nonna F. for always being there; to my Nonno G. for feeding me with his love and food, hope this makes 2019 slightly better. Thank you Delfi, for teaching me how to be a better me, we met almost at the end of my Ph.D journey, but we started a more important one. Crazy journey.

Last to thank, my Nonna M.. This is for you, *with love.*

Contents

List of Figures

List of Tables

Introduction

The controversial question of whether financial returns and equity premia are predictable has always attracted an enormous amount of attention, becoming one of the most debated topics in finance. The reason is that return predictability has not only a number of implications for financial models of risk and returns but also it is of interest to the daily work of practitioners. Given the central role of the predictability of stock returns, it is not surprising that the topic has attracted the attention of several authors along the years: Samuelson (1965), Fama (1970), Campbell (1987) Lo and Mackinlay, (1988) Shiller (2000) and Welch and Goyal (2007) among others. The idea of predictability of stock returns was initially interpreted as a rejection to the efficient market hypothesis by Samuelson (1965) and Fama (1970) and met with the empirical scepticism of the practitioners who gained abnormal returns. The evidence of such predictability brought several researches not only to test the degree of efficiency of the market but also to propose a plethora of approaches. These enhancements led to an increase of forecasting accuracy of financial returns: dividend-price ratio, the earnings price ratio, and the book-to-market ratio, size effect, book to market momentum effect (Lettau and Ludvigson (2001_a, 2004), Lettau and Nieuwerburgh (2008) and Welch and Goyal (2008) among others). Instead, another strand of the literature focused on investigating predictable characteristics of financial returns, such as the volatility to predict the evolution of the assets. Much of this has started from the development of the univariate GARCH class of models, Engle (1982) and Bollerslev (1986). Recently, this strand of research has benefited from the widespread of the high frequency data that allowed to achieve a direct measure of volatility, named Realized Volatility (Andersen and Bollerslev, 1998) Andersen et al. 2001, Barndorff-Nielsen and Shephard 2002, 2004) and to exploit the presence of long memory of the process for its prediction. Stock returns predictability has also been approached by proposing more sophisticated econometric techniques able to capture the market peculiarities, such as widespread evidence of time variation in the risk profiles of the assets and the presence of permanent abrupt exogenous shifts in the factor risk loadings (also named 'structural change'), which collectively may be described as instabilities in financial models. Such instabilities might be triggered by endogenous processes as in smooth transition models or probabilistic shifts between discrete regimes as in Markov switching models; or smooth parameterised stochastic time series processes, often random walks (Ferson and Harvey (1999), Adrian and Franzoni (2009); Gagliardini et al. (2015)). In practice, econometric models often incorporate parameter variation allowing random time varying coefficient models, driven by persistent random walk processes, bounded to avoid explosive outcomes. An alternative to the aforementioned approach is to use kernel methods to estimate time variation for deterministic parameter. Giraitis et al. (2014) developed an estimation method for random coefficient models using a kernel-based nonparametric technique as an alternative to state-space methods. They demonstrate that with only mild conditions on model parameters, the method has good statistical properties such as consistency and asymptotic normality in a range of finance-relevant contexts and settings. This thesis aims to contribute to the existing literature on asset predictability in 3 different ways.

In the first chapter, we make clear the importance of the investigation of financial volatility for a good prediction of stock returns. We propose then a comprehensive approach for capturing all the principal volatility features: the slow decay of the autocorrelation function, the asymmetric behaviour between volatility and asset returns and all the facets of volatility jumps, such as sign, magnitude and probability of occurrence. The contribution of this chapter is the fact that while a great deal of the work has produced improvements in the predictive accuracy of the financial volatility via a partial description of the phenomenon; no-one has yet considered all these volatility characteristics simultaneously in a single model. Recently, this strand of literature has benefited from the widespread use of high frequency data, that allowed computing a model free measurement of volatility, named Realized Volatility, RV, (by Andersen and Bollerslev, (1998); Andersen et al. (2001); and Barndor-Nielsen and Shephard, (2003)). Based on the properties of RV, Corsi (2009) proposed a simple but powerful model to capture the strong persistence in the autocorrelation function of the financial volatility series, known as long memory effect. He suggested considering the persistency in the volatility as the result of the interaction of the heterogeneity of agents behaviour over distinct time horizons and it employs a cascade of volatility components, one for each investments horizons to reproduce the dynamic that has the long memory property, with his Heterogeneous Autoregressive Realized Volatility model (HAR). Using this approach, different authors made an effort to enrich the description of the volatility phenomenon (Corsi et al. (2012), Clements and Liao (2013), and Patton and Sheppard (2015) among others) focusing on specific volatility features such as: leverage effect, jump magnitude and jump intensity, producing a limited description of the phenomenon. Our model, named, Leverage Heterogeneous Autoregressive Continuous, Jump, Intensity model, LHARCJI, extends Corsi's approach shedding light on all the volatility characteristics simultaneously. The empirical analysis, based on this model, not only produces an overwhelming evidence of the need to consider simultaneously all the features of volatility but also shows an interesting relation between level of jumps and the selected model. We find that the most predictive power of our approach is displayed in those assets which are more affected by jumps, such as Industrials, IT and Financials, where we observe, on average, around 5%-8% reduction in the loss functions across the assets, with peaks of reduction of 10%.

In the second Chapter, we presents a new hierarchical methodology for the estimation of a dynamic asset pricing model, employing an affine pricing kernel that does not require any *a priori* structure on the model. Despite the overwhelming evidence that risk premia varies over time (Campbell and Shiller, 1988; Cochrane, 2011), several asset pricing methods rely on the assumption that the price of risk is constant over time, e.g. Fama and MacBeth (1973). Assuming a smooth evolution of the risk premia over time, we decided to capture these movements using a kernel weighted repression approach. Hence, the main contribution of this Chapter is to introduce a hierarchical procedure, based on Fama and MacBeth (1973) model, for allowing a generic time variation in the risk premia and risk factor loading coefficients, using an extension of the kernel weighted regression specification proposed by Giraitis et al. (2014/2015), which can be seen as an extension of the frequently used least squares rolling window regression approach (Jagannathan and Wang (1996), Lewellen and Nagel (2006)).

Despite several approaches have been proposed to capture such time variation (Lettau and Ludvigson (2001_b) and Adrian and Franzoni (2009) and Adrian, Crump and Moench (2015)), the great advance of our methodology is that it avoids imposing any *a priori* structure and allows a natural data orientated way for incorporating economic and financial news that are relevant for the pricing of assets under investigation. We employ a flexible method for the bandwidth selection, which essentially determines the lag of recent updates of the betas (risk factors) and also of the factor loadings, identifying an optimal time varying bandwidth level optimised for each asset. The empirical results show that time variation of stocks and portfolios must be captured with an estimation tool that avoids imposing excessive *a priori* structure on the model while taking into account the specific features of each asset. Further, our methodological configuration allows a clear identification of the time varying relevance of each factor, being able, potentially to become, the starting point for a review of the literature on the factor loadings. Finally, our approach also shows improvements in forecasting performances of around 4%-7% in the one step ahead risk premium setting with respect to a plethora of alternative methods; independently the use of single stocks or portfolios. In Chapter 3, by applying an analytical approach, we investigate the temporal dependence in the autocorrelation functions shown by the volatility of a wide range of exchange rates and financial assets. Using the attractive features of the Realized Volatility measure that, among other things reduces emphasis of the type and choice of the model and enables a direct measurement of volatility, we examine the relationship between various long memory methodologies, such as fractionally integrated long memory models (see Baillie, 1996; Lamoreaux and Lastrapes 1990; Diebold, 1986), the Heterogeneous Autoregressive (HAR) model, by Corsi (2009) and its extended versions (Patton and Sheppard, 2015). It has been discovered that the Realized Volatility time series are characterized by strong persistence in their autocorrelations for a wide range of financial assets but no one has yet provided an explanation for this phenomenon, assessing whether it could be due to jumps, structural breaks, omitted non-linearities, contemporaneous aggregation or to pure long memory. We estimate HAR models generated by simulated fractional white noise processes and find that the simulated estimates have certain similarities with the HAR estimates from actual Realized Volatility data. We also estimate by maximum likelihood a restricted long memory model denoted by RARFIMA, which allows the long memory feature and also embodies parameter restrictions of the HAR model. The overall conclusion is that in many cases both the long memory feature and the HAR structure for short and medium term memory, can be important tools in representing variation within RV series. Finally, in accordance with the methodology introduced in Chapter 2, we also consider a kernel regression approach with constant bandwidth to estimate a time varying parameter HAR process. This model clearly shows that the relative importance of the partial volatility cascades varies throughout the sample. The model is quite effective in representing some of the long memory characteristics of Realized Volatility time series. However, the model selection results based on information criteria generally favour a simpler restricted ARFIMA structure with constant long memory and HAR parameters.

Outline of the Thesis

Chapter 1 introduces the approach to capture in a single linear model the volatility of assets returns. Once presented a review of the main characteristics of the Realized Volatility measures and its extended version, we introduce the theoretical framework of our approach in light of the seminal paper by Corsi (2009), highlighting the different methodologies to capture all the principal characteristic of the volatility. Finally, we report the empirical results of our out-of the-sample forecasting analysis aimed to identify the best model to predict one period ahead financial volatility.
Chapter 2 provides the details of the hierarchical methodology based on the identification of an optimal bandwidth for the estimation of a kernel weighted regression in the multi factor dynamic pricing setting, calibrated for different assets at different point in time. Once analysed the kernel methodology and the role of the bandwidth, we outline the cross validation procedure adopted to identify the optimal value for the bandwidth. Finally, the improvements of the methodology are validated using an out of sample forecasting exercise and a battery of robustness checks.
Chapter 3 investigates the importance of long memory parameters in the framework delineated by the Heterogeneous Autocorrelation model, Corsi (2009). We initially present details of the statistical quantities regularly implemented and arising from Realized Volatility series. Then, we describe the long memory models and inferential methods, report MLE of both ARFIMA models and also report semi parametric estimation of the long memory parameter. Further, simulation evidence is provided in support of the properties of OLS estimation of HAR models when the true data generating process is a fractional long memory process. Finally, we discuss the model performances in light of different asset classes.

Chapter 1

A new empirical approach for modelling stock market volatility

1.1 Summary

Understanding financial volatility has been one of the most debated strands in financial literature over the last decades. Different approaches have been proposed but, as far as we know, a comprehensive methodology has not been suggested yet. This paper introduces a new general approach for capturing in a single linear model all the keys features of the financial volatility. Our model extends the Heterogeneous Autoregressive model, Corsi (2009), and is able to capture not only the slow decay of the autocorrelation function but also the asymmetric behaviour between volatility and asset returns and all the facets of volatility jumps, such as sign, magnitude and probability of occurrence. Hence, we provide a more rigorous and comprehensive financial volatility model, able to improve the accuracy of the forecasting estimates for different assets. Indeed, an empirical analysis produced using fifty Standard & Poor's 500 constituents, overwhelmingly indicates the importance of considering all the aforementioned volatility characteristics in one unique model. The superiority of our methodology becomes evident especially for those assets more affected by jumps, such as Industrials, IT and Financials, where the improvements in the forecast accuracies are around 5%-8%, on average with peak of 10%.

1.2 Introduction

Modelling and forecasting volatility of stock market returns has always played a central role in many financial applications, e.g. modern pricing models, risk management theories and portfolio allocation. Recently, benefiting from the increase in the availability of the high frequency data, a new model-free measurement of volatility, called Realized Volatility, RV, has been suggested and computed by Andersen and Bollerslev (1998), Andersen *et al.* (2001) and Barndorff-Nielsen and Shephard (2003). Contrary to the classical approaches, such as (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, class of models proposed by Engle (1982) and Bollerslev (1986) and Stochastic Volatility (SV) models by Taylor (1986), this new methodology provides direct an observable proxy of financial volatility. Given such advantages, the Realised Volatility has been widely studied and used in the literature to improve the description of the volatility movements.

Among these attempts, Corsi (2009) proposed a very simple but powerful model to capture the strong persistence in the autocorrelation function of the financial volatility series, known as long memory effect. Stemming from the idea that the heterogeneity of agent′s behaviour over distinct time horizons affects differently the future volatility, he suggested to model the persistence using the past volatilities at different time horizons, with his Heterogeneous Autoregressive Realized Volatility model (HARRV). Thereafter, the simplicity of its structure and estimation, encouraged its use in several econometric studies. Corsi and Reno (2009), exploiting the differences in the asymptotic convergence of RV and Bi-Power Variation, BPV, included the predictive power of another volatility characteristics: the jump components. Corsi *et al.* (2012) model captured the asymmetric behaviour between volatility and return, known as leverage effect, extending the heterogeneous behaviour over different horizons to the past negative returns. Clements and Liao (2013), instead, focused their attention on the probability of jump occurrence, modelling it with a point process. Then, using the concept of semi variance, Patton and Sheppard (2015) showed that future volatility is more related with past negative returns rather than positive ones and that the impact of a price jump on volatility depends on the sign of the jump, with negative (positive) jumps leading to higher (lower) future volatility. Louzis *et al.* (2012) proposed to modify the way of modelling volatility to account asymmetries, considering simultaneously the absolute and standardised past returns at different horizons.

As an additional contribution to such literature, we observed that while a great deal of the work found indication for increasing the predictive accuracy of the models from a partial description of the phenomenon, no one has considered yet all these volatility characteristics simultaneously in a unique model. Therefore, the analysis conducted in this work aims to contribute to this literature, presenting a more comprehensive approach that combines at once all the volatility features. The model accounts for the long memory effect, leverage effect and for all the jump components (sign, magnitude and the probability of their occurrence): Leverage Heterogeneous Autoregressive Continuous, Jump, Intensity model, LHARCJI. A key application of this new estimator of financial volatility is in forecasting, since better measures of volatility enable us to better gauge the current level of volatility and better understand its dynamic, both of which lead to better forecasts of future volatility. Rejecting *a priori* the idea that a single model can fit all the scenarios, we test our specification on different asset classes in the S&P500 and against a plethora of competing models. In other words, we investigate the optimal model proposing a meticulous forecasting analysis fulfilled employing two different predictive ability tests: the Model Confidence Set (MCS) procedure recently developed by Hansen *et al.* (2011) and the pairwise forecasting test proposed in Giacomini and White (2006).

Our empirical results overwhelmingly indicate the importance of considering all the volatility features at once. The full superiority of our methodology becomes evident for those assets more affected by jumps, such as Industrials, IT and Financials, where we have a statistically significant reduction in the loss function of around 5%-8%, on average across the assets with peaks of around 10%. The model appears to be the best method in 7 out of 10 sectors with a rate of success, within them, of 90%.

The remaining part of this Chapter is structured as follows. Firstly, Section 1.3 provides a discussion of the contribution of this paper in light of the existing literature. Then, we outline the volatility notion describing its main features and provide the general univariate framework for introducing the Realized Volatility concept, Section 1.4. Hence, once the main characteristics of the latter are provided, we present our method for modelling the evolution of volatility, in light of the initial HAR model. Section 1.5 shows some techniques for handling the micro-structure noise in the construction of the realized volatility estimator and presents the empirical features of the assets under analysis. Finally, in Section 1.6, we display the empirical results of our out-of-sample forecasting analysis aimed at identifying the best model to predict one period ahead financial volatility.

1.3 Background literature

Over recent decades, understanding and modelling the volatility of asset returns has been one of the paramount strands in financial literature. The importance of volatility comes from the evidence that returns of financial assets are difficult to predict, whilst their volatility seems to show more exploitable features. However, an inherent problem of working with volatility is that it cannot be directly observed. Several attempts have been made to overcame this issue, starting from the seminal univariate (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, class of models proposed by Engle (1982) and Bollerslev (1986) and Stochastic Volatility (SV) models by Taylor (1986).

Although these classical volatility models and their developments provide nowadays very useful tools for volatility analysis, several authors have underlined that most of them are inadequate to depict all the stylised facts that characterise the financial time series, see e.g. Bollerslev (1987), Malmsten and Starica (2004) and Carnero *et al.* (2004). Specifically, they constantly fail to capture a key feature: the slow decay of the autocorrelation function, see Taylor (1986) and Ding *et al.* (1992). Also known as long memory property, it is considered one of the main causes for the volatility persistence. This feature has attracted much interest over the years and several solutions have been proposed to model it. The first attempt was made by Baillie *et al.* (1996) by introducing the Fractionally Integrated GARCH model, known as FIGARCH. Arising from the modification of a IGARCH model, its capacity of modelling intermediate range of persistence and its flexibility, make it one of the most important tools for capturing the long memory effect. Further attempts have been made focusing on the explanation of the causes of the long memory. Diebold (1986), Lamoureux and Lastrapes (1990), Mikosch and Starica (2004) and Hillebrand (2005) were among the first who argued that ignoring structural breaks might induce to a strong persistence in the autocorrelation function, explaining partially the long memory effects in financial time series. Several different tests were proposed to identify structural breaks in financial time series and reduce their impact on the persistency of the autocorrelation function, (among the others Kokoszka and Leipus, 2000; Andreou and Ghysels, 2002, 2006; Pesaran *et al.* 2013). Nevertheless, these remarkable enhancements still fail to overcame the main problem of the models where volatility plays a central role: the conditional variance is a latent variable, and hence must be computed. In recent years, the search of an adequate approach for volatility estimation and prediction has benefited from the considerable amount of high frequency intra-day data on equity and currency markets that has become available. This recent availability of data has led to the development of a growing strand of literature aimed to find a model-free measurement of variance (see McAleer and Medeiros, 2008 and Andersen *et al.*, 2002). First, Merton (1980) observed that, using high sampling frequency, the variance of a process over a fixed interval of time can be estimated by the sum of its squared realisations. More recently, following that idea, Andersen and Bollerslev (1997), Andersen *et al.* (2000, 2003), Barndorff-Nielsen and Shephard (2002, 2004), developed the concept of Realized Volatility, RV, which is a non-parametric ex-post measure of volatility at a daily frequency, obtained by the aggregation of squared intra-day returns, over a fixed time interval. This approach has the potential attraction of being an observed measure of volatility, which does not require the use of any underlying models, as opposed to GARCH and SV approaches.

The validity of this method hinges on an ex-ante analysis of the salient features of high frequency data, as suggested by Andersen *et al.* (2003), Hansen and Lunde (2006 a,b) and Brownless and Gallo (2006). Containing high amount of information, these type of data are exposed to the risk of market micro-structure noise (such as bid-ask bounce, asynchronous trading, missing values and infrequent trading) that might generate a divergence between the observed and the true value. Barndoff-Nielsen and Shepard (2003) showed that only in a perfect world where prices are monitored continuously and without noise, the RV converges to the Integrated Volatility, IV. Ait-Sahalia *et al.* (2006) and Bandi and Russel (2006, a, b) proved that RV has a reduced explanational power in presence of micro structure noise, converging to another quantity, named Quadratic Variation, which is the sum of IV plus a discrete jump component. The problem tends to worsen as the sampling frequency increases, underlying the existence of a trade-off between frequency and accuracy of the measure. Two different ways have been suggested to deal with this problem. Andersen *et al.* (2003) Oomen (2005), Hansen and Lunde (2006) proposed to reduce the bias using different sampling schemes, while Barndorff Nielsen and Shephard (2004, 2006) and Ait-Sahalia *et al.*

(2006), proposed to adjust directly the RV according different time-scale.
Subsequent work on Realized Volatility have generated a very large strand of literature aimed at describing volatility in a more accurate way than the classical approaches. Particularly, long-range dependence, leverage effects and the jumps component can be captured by a parsimonious linear structure model proposed by Corsi (2009), named Heterogeneous Autoregressive Realized model, HAR, and by its implementations. The model considers volatility persistence as the result of the interaction of the heterogeneous agent's perceptions of different time horizons, and employs a cascade of volatility components to reproduce the dynamic of the long memory property. Its simplicity and its easy economic interpretation have made the HAR models a very popular approach for modelling volatility in finance. Several extensions have been proposed, Corsi *et al.* (2009) suggested to include a GARCH structure for the HAR residuals to achieve more flexibility. Corsi and Reno (2009), employing different behaviour of volatility measure such as RV and BPV showed by Barndorff-Nielsen and Shephard (2007), divided the volatility in two parts, continuous and jump components. They found that most of the predictability arises from the continuous part, while jump process does not have any forecasting power. On the contrary, Clements and Liao (2013) also including in their model an estimate for the jump intensity, discovered that it contributes significantly to the improvements of forecasting accuracy measures. Then, Corsi *et al.* (2012), captured the leverage effect of the volatility, adding past negative returns at different time resolutions as regressors. Recently, Louzis *et al.* (2012) have proposed to account the asymmetric behaviour of returns and volatility and the long term structure through the inclusion of standardised and absolute returns aggregated over different periods, such as EGARCH model, and a FIGARCH approach for the conditional heteroskedasticity of the residuals. Finally, Patton ad Sheppard (2015), relying on the concept of semi-variance introduced by Barndorff-Nielsen *et al.* (2010), investigated the impact of signed returns on future volatility. They showed that future volatility is more related to negative realised semi-variance than positive one and disentangling the effects of these two components significantly improves volatility forecasts. Further, they obtained a measure of signed variation finding that the impact of a price jump on volatility depends on the sign of the jump, with negative (positive) jumps leading to higher (lower) future volatility.
In this study, we contribute to the existing literature by developing a new unified approach, aimed to increase the quality of the description and the prediction of the

financial volatility. We introduce a model that in addition to the Corsi's HARRV approach is able to capture simultaneously not only the slow decay of the autocorrelation function but also the asymmetric behaviour of volatility with respect to asset returns and the presence of jumps in all its components (sign, magnitude and probability of occurrence). Such approach also allows us to produce a detailed volatility analysis sector by sector, identifying which characteristics play a central role in the choice of the model specification of volatility.

1.4 Financial volatility

Volatility is one of the key elements in finance, playing a central role in many practical applications such as modern pricing models, risk management theories and portfolio allocation. For this reason, extensive efforts have been made to provide good real-time description of volatility.

Interpreted as a statistical dispersion measure of asset returns and naively computed as their standard deviation over a given period of time, it quantifies the amount of risk connected with the changes in the value of an asset. Generally, a higher volatility implies that a security's values can potentially vary significantly over a short period of time in either direction. Therefore, an accurate analysis of volatility is essential for investment purposes; since higher volatility implies higher profit but also higher risk.

The main issue is that volatility of returns is not directly observable. One of the approaches proposed in literature is to conduct inference using strong parametric assumptions, e.g. (G)ARCH and SV models. The worthiness of these is that the features of returns are well fitted by them, nevertheless they are affected by misspecification problems. An alternative is to employ option pricing models and use current market prices to deduce the volatility of the underlying instrument over a fixed future horizon, e.g. Implied Volatility. The main drawback is that such approach does not provide an unbiased estimator of volatility since it may incorporate a potentially time-varying risk premium. The third alternative to deal with the bias problem relies on *historical volatility* measures, which take into account the actual asset prices in the past. Among these techniques, the Realized Volatility overcomes the problem of bias offering a non-parametric ex-post measure of volatility at a daily frequency. The main advantage of the RV method is that, although it does not require any specification of any underlying models, it well depicts all the main stylised facts of volatility series: mean reversion, fat tails, long range dependence, jumps.

1.4.1 Realized Volatility

Realized Volatility is a model free measurement of financial volatility, proposed by Andersen and Bollerslev (1998), Andersen *et al.* (2003), Barndorff-Nielsen *et al.* (2003). The intuition behind it comes from the popular continuous-time diffusion process, characterised by the absence of jumps:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \quad t \geq 0, \tag{1.1}$$

where $dp(t)$ is the logarithm price increment at time t, $\mu(t)$ denotes the drift term that has continuous and locally bounded variations, $\sigma(t)$ is a strictly positive spot volatility process and $W(t)$ is a standard Brownian motion. Assuming that the time length of a day is 1, the daily-returns can be written as

$$r_t = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s). \tag{1.2}$$

We observe from Equation (1.2) that the daily volatility of stock returns is related to the evolution of the spot volatility, σ_t. Thus, the distribution of the stock returns depends on both the drift and spot volatility components, $r_t \sim N\left(\int_{t-1}^{t} \mu(s)ds, \int_{t-1}^{t} \sigma^2(s)\right)$ where the second term is called Integrated Volatility variation, IV. The corresponding discrete-time within-day return is

$$r_{t+j\Delta} = p(t + j/M) - p(t + (j-1)/M), \qquad j = 1, 2, \ldots, M,$$

where M denotes the number of intra-day, equally spaced, returns over t, and $\Delta = 1/M$ refers to the sampling interval. In such framework, the Realized Volatility, RV, is defined as the sum of equally spaced high frequency squared returns, over a fixed time interval:

$$RV_t(\Delta) = \sum_{j=1}^{M} r_{t+j\Delta}^2. \tag{1.3}$$

Andersen *et al.* (2003) showed that under suitable conditions (such as absence of serial correlation in the intra-day returns and absence of micro-structure noise), the RV is a consistent estimator of the IV,

$$lim_{M\to\infty} RV_t(\Delta) = \int_{t-1}^{t} \sigma^2(s)ds = IV_t. \tag{1.4}$$

Despite the great success of the Realized Volatility measure, it has also been acknowledged that, practically, high frequency returns are affected by microstructure noise[1] and jumps that might introduce autocorrelations among the observations and alter the asymptotic characteristics of the RV. Although these effects can be reduced using an adequate sample frequency, it is necessary to take into account the jump process, for a more realistic analysis.

Suppose the log-price process is described using a Brownian Semi-Martingale with Jumps family models:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \quad t \geq 0. \tag{1.5}$$

where $\kappa(t)dq(t)$ describes the jump component, with $\kappa(t)$ size of the jump, $q(t)$ a continuous process with $q(t) = 1$ if there is a jump at time t and 0 otherwise and a hidden parameter, called jump intensity ($\lambda(t)$), which represents the probability of jump occurrence (in a discrete representation jumps follow a Poisson process). This representation implies that the continuous diffusion process captures the smooth variations of the log-prices, while on the other hand the jump component accounts for discontinuities. The corresponding discrete-time daily returns are defined as

$$r_t = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma^2(s)dW(s) + \sum_{s=1}^{N(t)} \kappa^2(s),$$

where $N(t)$ counts the number of jumps occurring with possible time varying intensity. Andersen *et al.* (2007) have shown that in presence of jumps, RV converges uniformly

[1]These effects can be reduced using an adequate sample frequency. Andersen *et al.* (1998) recommend to sample at a moderate frequency, such as 5 minute or 15 minute to optimise the trade-off between bias and amount of information. Alternatively, Zhang (2006), Ait-Sahalia *et al.* (2006) and Bandi and Russel (2006) suggested different computation approaches to reduce the bias, employing very high frequency. Nevertheless, Brownlees and Gallo (2006) showed that the risk of measurement error and endogeneity persist, see McAleer and Medeiros (2008) for more details.

in probability to a sum of Integrated Variance and jump process:

$$lim_{M\to\infty} RV_t(\Delta) = IV_t + \sum_{t\leq s\leq t-1} \kappa^2(s). \tag{1.6}$$

Hence, RV is a consistent estimator of the IV only in absence of jumps, while otherwise it converges to a quantity that accounts for the jump process, $\sum_{t-1\leq s\leq t} \kappa^2(s)$.
Several different approaches have been proposed for achieving a consistent estimation of IV, but for our work we focus only on the well-established tool proposed by Barndorff-Nielsen and Shephard (2004)[2]. They proved that, under the regular condition that jumps have finite activity, the normalised sum of products of the adjacent absolute values of returns, (i.e. Bi-Power Variation, BPV) is a consistent estimator of the IV also in presence of jumps. The BPV is defined as

$$BPV_t(\Delta) = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{j=2}^{M} |r_{t+j\Delta}||r_{t+(j-1)\Delta}|, \tag{1.7}$$

while its convergence to IV is shown by relation

$$lim_{M\to\infty} BV_{t+1}(\Delta) = \int_{t-1}^{t} \sigma^2(s)ds = IV_t. \tag{1.8}$$

BPV is designed to be robust to jumps being built on two consecutive absolute returns and not on squared returns as RV. Indeed, the presence of jumps is not amplified, as for the RV, but it is downsized by the other value giving a negligible impact of the jumps on the asymptotic behaviour of the BPV. Finally, Barndorff-Nielsen *et al.* (2010) introduced a measure, called Realized Semi variance (RS) for identifying the contribution of negative and positive returns to the volatility. The positive (negative) realized semivariance, RS^+ (RS^-), is computed by summing the squared intra-day

[2]Several different techniques, robust to microstructure noise and jump estimators of IV, were discussed in McAleer and Medeiros (2008) and Mancini and Calvori (2012). These techniques suggested to reduce the impact of jumps and microstruture noise in the asymptotic behaviour of RV using a composition of RV measure computed at different time horizon.

returns associated with positive (negative) returns:

$$RS_t^{+} = \sum_{i=1}^{M} r_{t,\tau}^2 I\left\{r_{t,\tau} > 0\right\}, \tag{1.9}$$

$$RS_t^{-} = \sum_{i=1}^{M} r_{t,\tau}^2 I\left\{r_{t,\tau} < 0\right\}, \tag{1.10}$$

where $I(\cdot)$ is an indicator function. These estimators provide a complete decomposition of the realized volatility, such that $RV_t = RS_t^{+} + RS_t^{-}$. Further, their limiting behaviour has been proved to be the combination of both the continuous part of volatility process, represented by the half of IV component and the jump component.

$$RS_t^{+} \underset{p}{\rightarrow} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} \kappa^2(s) I\left\{\kappa > 0\right\}, \tag{1.11}$$

$$RS_t^{-} \underset{p}{\rightarrow} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} \kappa^2(s) I\left\{\kappa < 0\right\}. \tag{1.12}$$

1.4.2 Modelling the jump component

Given the key role played by the jump process in the asymptotic behaviour of RV, and shown in Section 1.4.1, several approaches have been introduced to modelling the jumps in all their components. Barndorff-Nielsen and Shephard (2004) exploited the different asymptotic behaviour of RV and BPV in order to estimate the contribution of jump to volatility. Combining these two, it is possible to consistently compute the magnitude of jump process:

$$plim_{M\rightarrow\infty}\left(RV_t(\Delta) - BPV_t(\Delta)\right) \rightarrow \sum_{t-1<s<t} \kappa^2(s).$$

Further, since nothing prevents estimates of the squared jumps from becoming negatives in a given sample, the authors suggested truncating them at zero,

$$J_t(\Delta) = max\left[RV_t(\Delta) - BPV_t(\Delta), 0\right]. \tag{1.13}$$

Subsequently, in another article, Barndorff-Nielsen and Shephard (2010) noted that using the asymptotic behaviour of the Realized Semi-Variance, it is possible to extract a measure of the signed jump process. What remains after removing the variation due to the continuous component, subtracting one RS from the other, is called signed jump variation:

$$\Delta J^2 \equiv RS_t^+ - RS_t^- \underset{p}{\rightarrow} \sum_{t-1<s\leq t} \kappa^2(s) I\left\{\kappa > 0\right\} - \sum_{t-1<s\leq t} \kappa^2(s) I\left\{\kappa < 0\right\}, \qquad (1.14)$$

where $I(\cdot)$ denotes an indicator function. Following such approach, Patton and Shepard (2015), proposed to disentangle the negative and positive jump variation to prove that the impact of jumps is driven more by negative variation rather than positive one and that both have a negative sign. Such finding reveals that days dominated by negative jumps lead to higher volatility while days with positive jumps lead to lower volatility. They proposed to use an indicator function to identify when the difference in the realized semi-variance is greater or lower than zero, as a proxy for the sign of the jump process:

$$\Delta J^{2+} = RS_t^+ - RS_t^- I(RS_t^+ - RS_t^- > 0), \qquad (1.15)$$

$$\Delta J^{2-} = RS_t^+ - RS_t^- I(RS_t^+ - RS_t^- \leq 0), \qquad (1.16)$$

where $I\left(\cdot\right)$ is an indicator function. Andersen *et al.* (2007) criticised this approach, underlying two critical points. Firstly, they noted that its theoretical justification is based on the assumption of an infinite sample; hence, any finite sample implementation is subject to measurement errors. Secondly, they observed that although the truncation for non-negative values alleviates the previous type, this is not sufficient since jump estimates still may exhibit an unreasonable large number of very small positive values. Starting from such critique, a more sophisticated and powerful approach for the jump detection was developed by Barndorff-Nielsen and Shephard (2007). They proposed to treat these large number of very small positive jumps as measurement error, focusing only on significantly large jumps. Then, they developed an improved version of the ratio-statistic by Haung and Tauchen (2005) for the detection of jumps. The ratio-statistic is defined as follows:

$$Z_t(\Delta) = \Delta^{1/2} \frac{\left[RV_t\left(\Delta\right) - BV_t(\Delta)\right] RV_t\left(\Delta\right)^{-1}}{\left[\left(\frac{\pi^2}{4} + \pi - 5\right) \max(1, TQ_t(\Delta) BV_t(\Delta)^{-2})\right]^{1/2}}. \qquad (1.17)$$

where $TQ(\Delta) = \Delta^{-1}\mu_{4/3}^{-3}\sum_{j=3}^{1/\Delta}|r_{t+j\Delta,\Delta}|^{4/3}|r_{t+(j-1)\Delta,\delta}|^{4/3}|r_{t+(j-2)\Delta,\Delta}|^{4/3}$ is the Realized Tripower Quarticity[3]. Then, the jump and the continuous components were estimated by the realisation of $Z_{t+1}(\Delta)$ in excess of some critical value:

$$J_{t,\alpha}(\Delta) = I(Z_t(\Delta) > \Phi_\alpha) \cdot [RV_t(\Delta) - BV_t(\Delta)], \tag{1.18}$$

$$C_{t,\alpha}(\Delta) = I(Z_t(\Delta) \leq \Phi_\alpha) \cdot (RV_t(\Delta) + BV_t(\Delta)), \tag{1.19}$$

where $I(\cdot)$ denotes an indicator function, Φ_α is the value of the standard Normal distribution. The formulation ensures that the two components add to the original values of RV[4].

Recently, Clements and Liao (2013) focused their research interest to increase the description of jumps. They investigated not only the magnitude of the jumps but also the probability of occurrence, interpreting it as a point process[5]. Specifically, let $\{t_i\}_{i\in 1,\ldots,M}$ be a random sequence of increasing events $0 \geq t_1 > \cdots > t_M$, which describes a simple point process, and $M(t) := \sum_{i\geq 1}\mathbf{1}_{t_i\geq t}$ is a counting function. The conditional density, $\lambda(t)$, can be seen as the expected variation in $M(t)$ over a small interval,

$$\lambda(t) = \lim_{s\to t}\frac{1}{s-t}E\left[M(s) - M(t)\right]. \tag{1.20}$$

A common specification of $\lambda(t)$ is the self-exciting Hawkes process, Hawkes (1971), (for a more details see Bauwens and Hautsch, 2009):

$$\lambda(t) = \mu + \int_0^t w(t-u)dM(u) = \mu + \sum_{t_i<t} w(t-t_i), \tag{1.21}$$

with μ a constant and $w(\cdot)$ a non-negative decreasing weight function of $t-u$, implying that after a spike the intensity decreases. This is a self-exciting process since $Cov\left[N(a,b), N(b,c)\right] > 0$ where $0 > a \geq b < c$. Clements and Liao (2013) proposed a discrete version of Equation (1.21), where $\lambda(t)$ denotes the discrete time conditional

[3]Barndorff-Nielsen and Shephard (2007) model preserves the characteristic of the original statistic (17), in particular the approximation of Z_t to the standard normal distribution. The TQ_t is an estimator of the Integrated Quarticity, $IQ_t = \int_0^1 \sigma^4(t+\tau-1)d\tau$.

[4]Corsi *et al.* (2010) refined the above method suggesting to use the Threshold Bi-Power Variation, $TBPV$, to estimate the continuous part of the process instead of the BPV.

[5]The point process is a common approach for describing the arrival of trade and quotes, see Bauwens and Hautsch (2009).

intensity for the $t - th$ trading day. Specifically, it is computed using the past occurrences, to allow for self-excitation of modelling short persistence, and exogenous variables, such as the level of volatility:

$$\lambda_t = \mu + \gamma X_{t-1} + \begin{cases} \beta \quad \lambda_{t-1} & \text{if } dN_{t-1} = 0 \\ \beta \quad \lambda_{t-1} + \alpha & \text{otherwise} \end{cases} \tag{1.22}$$

where μ is the baseline intensity, λ_{t-1} is the intensity at $t-1$, β denotes the decay factor of the intensity and α is the shock to the intensity on day t if the jump is occurred at $t-1$. Further, X_{t-1} is the set of the exogenous variables; here, it includes the lags of BPV[6]. In conclusion Clements and Liao (2013) proved that the intensity of jumps is self-exciting with the level of intensity influenced by the level of volatility.

1.4.3 A unified model for financial volatility

Allowing a direct observation of financial volatility, RV has been exploited to provide an alternative to the classical GARCH and SV approaches. Among all the proposed approaches, HAR method, by Corsi's (2009), has attracted a huge amount of attention, because of its simplicity of computation, easy economic interpretation and flexibility. Introduced to capture the presence of long memory in volatility, it was subsequently extended to account also for the presence of jumps and asymmetric behaviour between volatility and negative returns.

The seminal Corsi's (2009) model is based on the observation that since volatility persistence can be caused by the heterogeneous agent's perception of time horizons, the volatilities over different time intervals might have different influences. Generally, the level of short term volatility does not affect the strategy of long term agent while, instead vice-versa is true. He proposed an additive linear model of heterogeneous RV components, one for each investment horizons, daily-weekly-monthly, to reproduce the dynamic of long memory effect and hence the persistence of the financial volatility. This lead to an AR-type approach in the RV, hereafter called *Heterogeneous Autoregressive*, HARRV. Considering the logarithmic transformation of RV to avoid negativity issues

[6]Clements and Liao (2013) showed that BPV and other jump robust measures of volatility such as the Threshold Bi-Power Variation , $TBPV$, produces statistically identical results. For this reason, we report in the main body of the paper only the results for BPV as exogenous variables. Results for other proxies are available upon request.

and get approximately Normal distributions, this model can be expressed as

$$\log RV_t^d = \alpha_0 + \alpha_1 \log RV_{t-1}^d + \alpha_2 \log RV_{t-1}^w + \alpha_3 \log RV_{t-1}^m + \varepsilon_t, \tag{1.23}$$

where $\log RV_t^d$ is the daily log-realized volatility, $\log RV_{t-1}^w = \sum_{i=1}^{5} \log RV_{t-i}$ denotes the weekly log-realized volatility and $\log RV_{t-1}^m = \sum_{i=1}^{22} \log RV_{t-i}$ represents the monthly log-realized volatility. The ε_t is an $i.i.d$ random variable with zero mean and unit variance, and coefficients are estimated by ordinary least square method with Newey-West covariance correction for serial correlation[7].

Moving from this framework, we propose a unified model for accounting simultaneously not only the strong persistence in the autocorrelation function but also other volatility features such as: the sign of the jumps (Patton and Shepard, 2015), their magnitude (Andersen *et al.*, 2007) and their probability of occurrence (Clements and Liao, 2013) and the leverage effect (in accordance with Corsi *et al.*, 2012). Extending the heterogeneous structure proposed by Corsi (2009) to all the components of volatility arrive at the following model :

$$\begin{aligned} \log RV_t^d = {} & \alpha_0 + \alpha_1 \log C_{t-1}^d + \alpha_2 \log C_{t-1}^w + \alpha_3 \log C_{t-1}^m \\ & + \alpha_4 \log(J_{t-1}^{d-} + 1) + \alpha_5 \log(J_{t-1}^{w-} + 1) + \alpha_6 \log(J_{t-1}^{m-} + 1) + \alpha_7 \lambda_t \\ & + \alpha_{11} r_t^{d-} + \alpha_{12} r_t^{w-} + \alpha_{13} r_t^{m-} + \epsilon_t, \end{aligned} \tag{1.24}$$

where $\log C_{t-1}^w = 1/5 \sum_{i=1}^{5} \log C_{t-i}$ and $\log C_{t-1}^m = 1/22 \sum_{i=1}^{22} \log C_{t-i}$ is the continuous volatility component computed according to Equation (1.19). Recognising the importance of including the sign of the variation and at the same time to avoid the inclusion of noise from the small jumps, we decided to define the negative jump components as follows:

$$J_{t,\alpha}^{-}(\Delta) = \{I(RS_t^{+} - RS_t^{-} < 0) \cup I(Z_t(\Delta) > \Phi_\alpha)\} \cdot [RV_t(\Delta) - BV_t(\Delta)]. \tag{1.25}$$

[7]The model is specified in logs according the original version in Corsi (2009) to avoid non-negativity issues and allow Normal approximation. The results reported in this paper are run using the logarithm transformation, but are robust to such transformation and available upon request.
An alternative method to the log transformation is to use the simple weighted least squares (WLS). To implement this, we first estimate the model using ordinary least square, then construct weights as the inverse of the fitted value from that model. Such approach is motivated by considering the residuals of the regression to have heteroskedasticity related to the level of the process. The results are available upon request, see Corsi (2009).

In other words, our approach is a combination of different techniques aimed at considering the different impact of only statistically significant jumps when the return is negative. The choice to exclude the positive component of jumps, and $J^{+}_{t,\alpha}(\Delta) = \{I(RS^{+}_t - RS^{-}_t > 0) \cup I(Z_t(\Delta) > \Phi_\alpha)\} \cdot [RV_t(\Delta) - BV_t(\Delta)]$, is motivated by the relative magnitude of the coefficients and limited significance of this variable as shown by Patton and Shepard (2015)[8]. The $\log(J^{w\cdot}_{t-1}+1) = \sum_{i=1}^{5} \log(J^{\cdot}_{t-i}+1)$ and $\log(J^{m\cdot}_{t-1}+1) = \sum_{i=1}^{22} \log(J^{\cdot}_{t-i}+1)$ are respectively the heterogeneous extensions of positive and negative jumps at different time horizon. Intuitively, C, represents the persistent volatility in the market due to the uncertainty about the future path of the returns discounted by the presence of sudden shocks in the market, i.e. jumps[9]. λ_t is a measure of jump activity, computed according Equation (1.22); specifically, it describes the probability of observing a jump today given a jump occurred yesterday, providing a synthetic measure of how much the process is affected by jump component. $r^{m-}_t = \max(r^{(n)}_t, 0)$ and $r^m_t = 1/n \sum_{i=1}^{n} r_{t-i}$ with $m \in [d; w; m]$ and $n \in [1, 5, 22]$ are lagged negative returns occurred at different time horizon that capture the asymmetric behaviour of volatility to previous daily, weekly and monthly negative returns. The innovation process, ϵ_t is an *i.i.d* random variable with zero mean and unit variance[10].

1.5 Properties of the data

Before introducing the main characteristics of our data, we present a brief introduction of high frequency data highlighting the main difficulties and how to minimise them.

High Frequency Data refers to a dataset containing detailed information of all financial market activities. In literature, they are also named tick-by-tick data emphasising its atomic unit of information. It can be seen as the transaction price or a order book that contains sell and buy orders at any point in time.

A high frequency database contains a huge number of ticks per day. Approximately, there are 2000 ticks per day per single stock (850000 per year), but this is only around estimate, due to the randomness related to the time interval which separates following

[8]Results for such unrestricted models are available upon request from the authors.

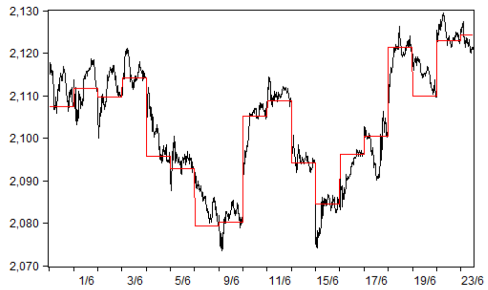
[9]Different approaches for disentangling the continuous and jump components have been used. Here, we report only the one proposed by Barndorff-Nielsen and Sheppard (2007) while the others have been used as robustness check and are available upon request.

[10]An analytical description of the competing models is provided in Appendix A

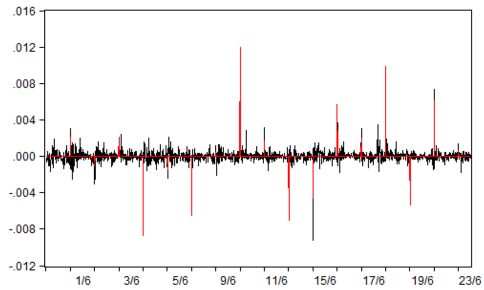
ticks (asynchronous trading). The sequence of ticks might also be characterised by missing values and/or exhibit some anomalies due to market conditions, named market micro structure noise. Engle (2000) and Bauwens and Giot (2001) showed that the sequence and the structure of ticks strongly depends on the collecting procedures and market mechanisms. Therefore, an ex-ante analysis of the data-structure is essential to remove all these potential sources of bias. Specifically, two preliminary steps must be taken into account before performing an econometric analysis: Cleaning and Management manipulation.

Cleaning operation – It concerns all the procedures aimed to detect wrong observations and those that do not reflect the market activity. The main sources of these types of errors are: human mistakes and trading intensity. Indeed, every second a high amount of data must be recorded and despite using a fully automated system, it might be that some data is recorded incorrectly. Hence, the idea behind this phase is to remove all those observations that are not coherent according to the chosen distance criteria (see for instance Brownlees and Gallo, 2006, and Zhou,1996).

Rescaling operation – The second phase is related to the construction of the series for analysis. The presence of market micro structure noise prevents the use of most of the models based on equally spaced time intervals. An appropriate aggregation function must be employed for irregular spaced high frequency data in order to have an equally spaced series and reduce the effect of the noises, as shown by Andersen *et al.*(2003). Some examples are: calendar time sampling (CTS), transaction time sampling (TrTS), and tick time sampling (TkTS), by respectively Oomen (2005), Hansen and Lunde (2006), McAleer and Medeiros (2007). Among these, the most used procedure is a variation of the TrTS, by Andersen *et al.* (2003). It optimises the trade-off between the effect of noise and amount of information used for the description of the volatility. It suggests to sample at a moderate frequency, such as 5 minute or 15 minute and in presence of missing values to use the previous tick price or an average of the previous and after ticks. The final outcome of this procedure is a sequence of high frequency data equally spaced, at a given frequency, in which all the observations are coherent with the market activity.



(a) Prices



(b) Returns

Note: The Figures displays a comparison between daily and equally spaced (five minute) intra-day prices for the S&P500 stock market index, on the left panel, while on the right one the comparison among the respective returns. The red line plots daily observations, while the black one plots the intra-day data (5 min). The sample spans from 29/05/2017 to 23/06/2017 and covers a trading day time defined as the period 9:30AM - 16:05PM.

FIGURE 1.1: Daily and intra-day (5-min sampling) data for the S&P500 index

1.5.1 Issues in handling intra day databases

Once analysed the aforementioned issues of employing the high frequency data, it is important to understand the improvements achieved in the description of both prices and returns, using such type of data changes. Figure 1.1 plots on the right panel a comparison between daily and 5-minute interpolated transaction prices for the S&P500 index, while on the left hand one the same comparison is made using their returns. The cleaned and adjusted sample spans from 29/05/2017 to 23/06/2017 and covers a trading day time defined as the period 9:30AM - 16:05PM, for a total of 17 working days and 1442 observations. All the data is available on Bloomberg.

The graphical analysis clearly shows the presence of different amounts of information inside of the different datasets. Despite daily prices were computed using high frequency data, they completely neglect intra-day movements plotting only a general tendency of the path. Conversely, the HFD well describes all oscillations, allowing to figure out that daily values are exclusively proxies for intra-day movements.
This type of information also affects the distributional properties of the returns, that vary with the sampling frequency. It has been shown that at higher frequency, there is

a stronger evidence of returns distribution being non-Gaussian, Table 1.1. Indeed the descriptive statistics for the high frequency returns, although having an approximately symmetric distribution and a finite second moment, they show a very high fourth moment, especially for 1-minute sampling. Such results show again the difficulties related with the use of HFD but also the perks of them.

Table 1.1: Summary statistics of different frequency S&P500 returns

	Mean	**St. Dev.**	**Min**	**Max**	**Skew**	**Kurt**
Daily	0.0006	0.0620	-0.8661	1.1971	-0.3219	43.35
30-min	0.0002	0.0504	-1.0987	0.9820	-0.4457	50.00
5-min	0.0005	0.0677	-0.9186	0.7331	-0.6712	107.62
1-min	0.0003	0.0624	-0.8186	0.9314	-0.8521	250.46

Note: The Table provides the descriptive statistics for the S&P500 returns computed at difference frequencies. The cleaned and adjusted sample spans from 29/05/2017 to 23/06/2017 and covers a trading day time defined as the period 9:30AM - 16:05PM, for a total of 17 working days and 1342 observations.

1.5.2 Statistical description of the data

For our analysis we consider 50 of the S&P500 constituents divided into 10 Sectors according the Global industry Classification Standard (GICS), Table 1.2. The choice of the following dataset is based on the transaction volume during the past 3 years (subject to data availability), and aimed at finding commonalities in the financial volatility within the GICS sectors. Our primary data consists of tick-by-tick transaction prices, ranging from January 30, 2002 to May 31, 2017, apart from the CBS Corp. asset that starts from 03/01/2005. The trading hours span from 9:30 AM to 16:00 PM with around 120 observations per hour. We adopt the method proposed by Andersen *et al.* (2007) for computing our Realized Volatility and jump measures, sampling the return and using the nearest prices to each five minute mark for the most actively traded contracts[11]. The choice to sample prices using an approximative five minute sampling

[11]The volatility signature plot for the same data depicted in Corsi *et al.* (2008) suggests that the returns are largely immune to the contaminating influences of the market micro structure noise at that frequency. In particular, the ratios of the sample means of the five-minute based realized measures to the ones based on 15 and 30-min sampling, equal 0.9936 and 0.9746 for the realized variance, and 0.9732 and 0.9660 for the Bi Power Variation, respectively.

TABLE 1.2: S&P500 constituents and GICS classification

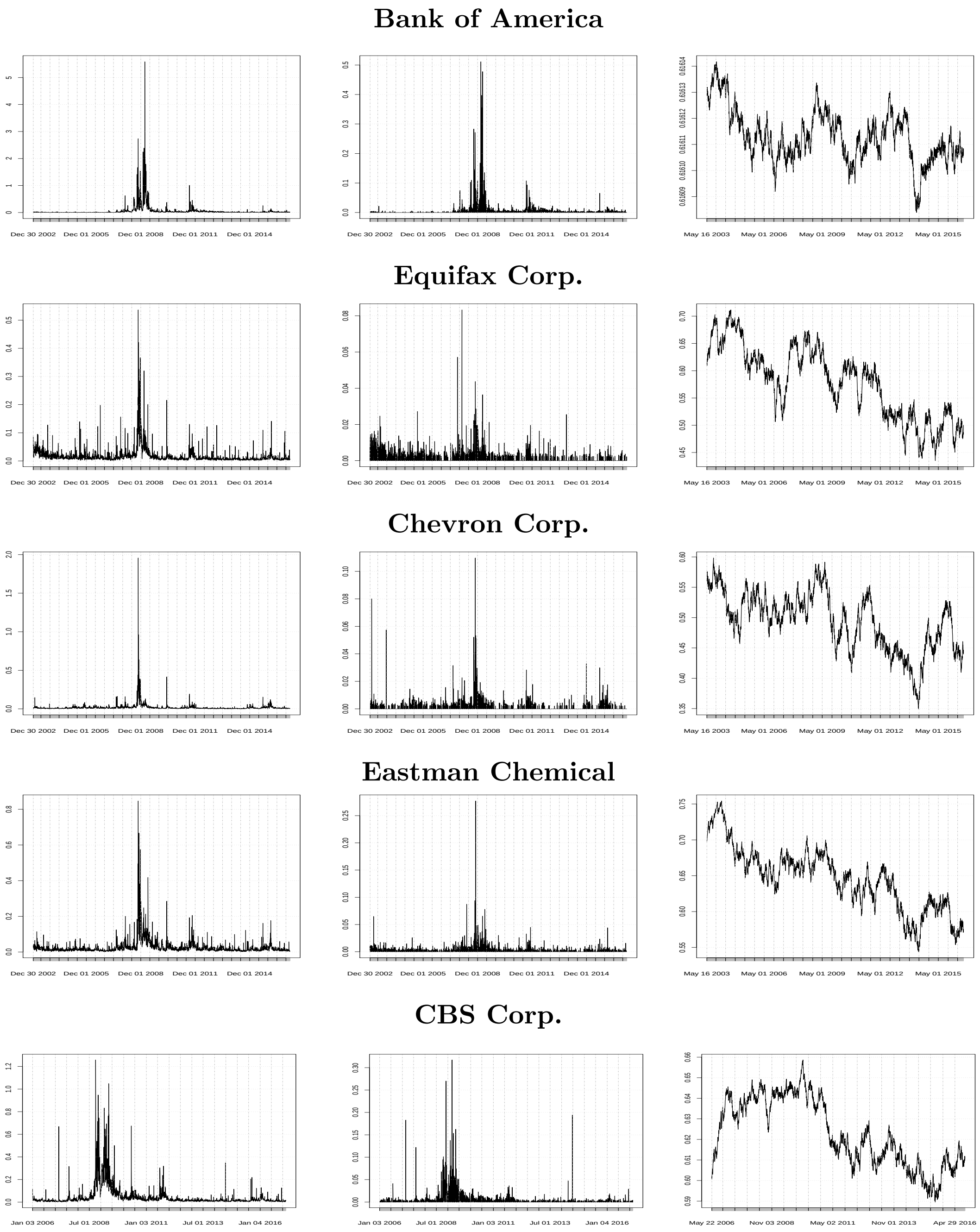
	Company name		Company name
Financial		**Health**	
JPM	JPMorgan Chase & Co.	BAX	Baxter Int. Inc.
BLK	BlackRock, Inc.	ABT	Abbott Laboratories
BAC	Bank of America Corp	JNJ	Johnson & Johnson
AXP	American Express Comp.	MDT	Medtronic PLC
WFC	Wells Fargo	PFE	Pfizer Inc.
Industrial		**IT**	
ALK	Alaska Air Group Inc	AAPL	Apple Inc.
EFX	Equifax Inc.	EBAY	eBay Inc.
FDX	FedEx Corp.	AMZN	Amazon.com, Inc.
UNP	Union Pacific	INTC	Intel Corp.
KSU	Kansas City Southern	ADBE	Adobe Systems Inc.
Energy		**Utilities**	
XOM	Exxon Mobil Corp.	DUK	Duke Energy
PXD	Pioneer Natural Resources	AEP	American Electric Power
CVX	Chevron Corp.	PPL	PPL Corp.
APA	Apache Corp.	FE	FirstEnergy Corp
SLB	Schlumberger Ltd.	ED	Consolidated Edison
Materials		**Real Estate**	
VMC	Vulcan Materials	BXP	Boston Properties, Inc.
WRK	WestRock	FRT	Federal Realty Inv. Trust
AVY	Avery Dennison Corp	VTR	Ventas Inc
APD	Air Products	PSA	Public Storage
EMN	Eastman Chemical	SPG	Simon Property Group
Cons. Discret.		**Cons. Staples**	
RL	Polo Ralph Lauren Corp.	PM	Philip Morris Int.
DIS	The Walt Disney Comp.	CVS	CVS Health
TSCO	Tractor Supply Comp.	K	Kellogg Comp.
RCL	Royal Caribbean Cruises Ltd.	CL	Colgate-Palmolive
CBS	CBS Corp.	KO	Coca-Cola Comp.

Note: Table provides the list of the assets under analysis divided into 8 Sectors according the Global industry Classification Standard (GICS). The datasets are obtained from Alpha Trading, (alphatrading.com).

period is a standard one and is motivated by the desire to avoid bid-ask bounce type microstructure noise. This sampling schemes let us with 78 intra-day observations for a total number of around 295000 observations for each asset. Then, we produce the volatility estimates, using ten different grids of the equally spaced prices to obtain ten different estimators, which are correlated but not identical. At the end, we average

these to obtain our final estimator. Such an approach proposed by Zhang *et al.* (2005), aims to produce a mild increase in precision relative to using a single estimator.
The resulting daily series for realized volatility, jumps and intensity of jumps are displayed in Figures 1.2 and 1.3 for a representative asset for each GICS Sectors and computed respectively according to Equations (1.3), (1.18) and (1.22)[12]. The assets are: Bank of America, Equifax Corp., Chevron Corp., Eastman Chemical, CBS Corp., Baxter International, eBay Inc., Duke energy, Boston Properties Inc., Philip Morris International.

[12]A wider view about the characteristics of the data and of the volatility measure can be found in Appendix B, where we report the analysis for each of the asset under consideration.

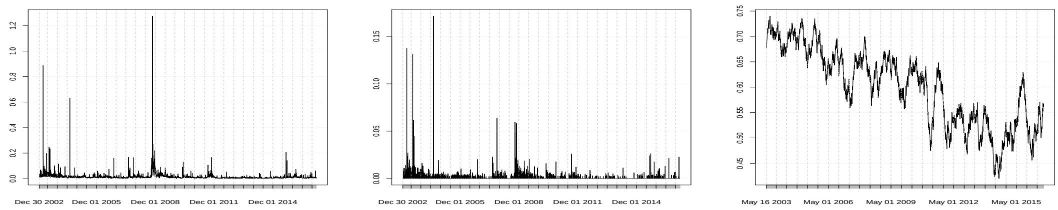


(a) RV (b) Jump (c) Intensity

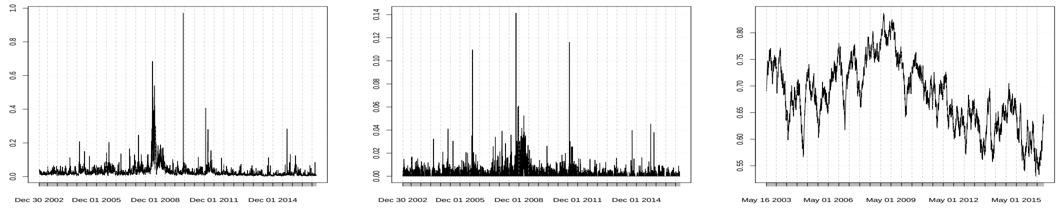
Note: The figure provides the plots for Realized Volatility, Jumps and Intensity of Jumps for representative assets for each Sector. The RV series are computed according Equation (1.3); the jump series instead are computed using Equation (1.18), while the jump intensities according Equation (1.22), where we assumed the explanatory variable is BPV.

FIGURE 1.2: Realized Volatility, Jump, and Intensity by Sector

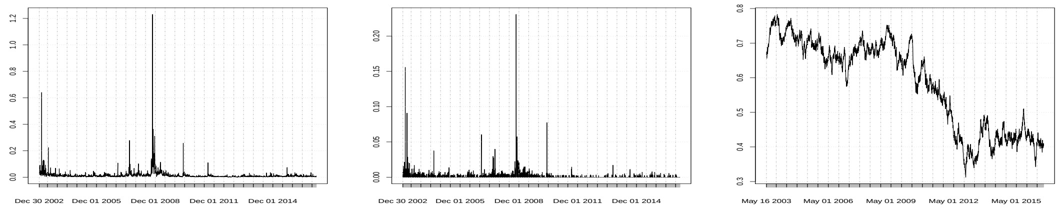
Baxter International Inc.



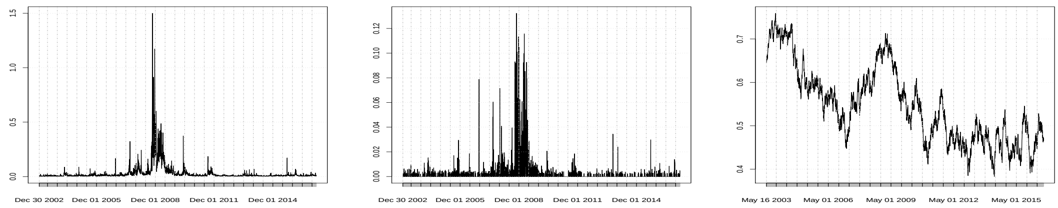
eBay Inc.



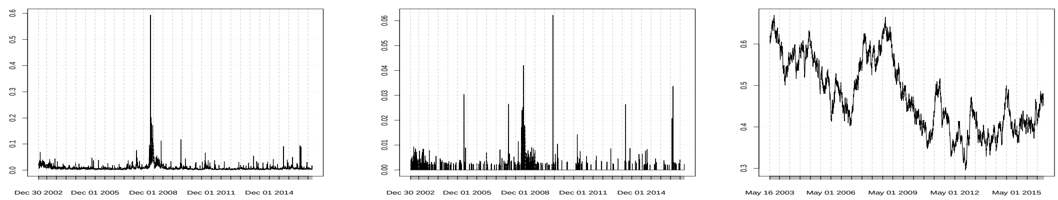
Duke Energy



Boston Properties Inc.



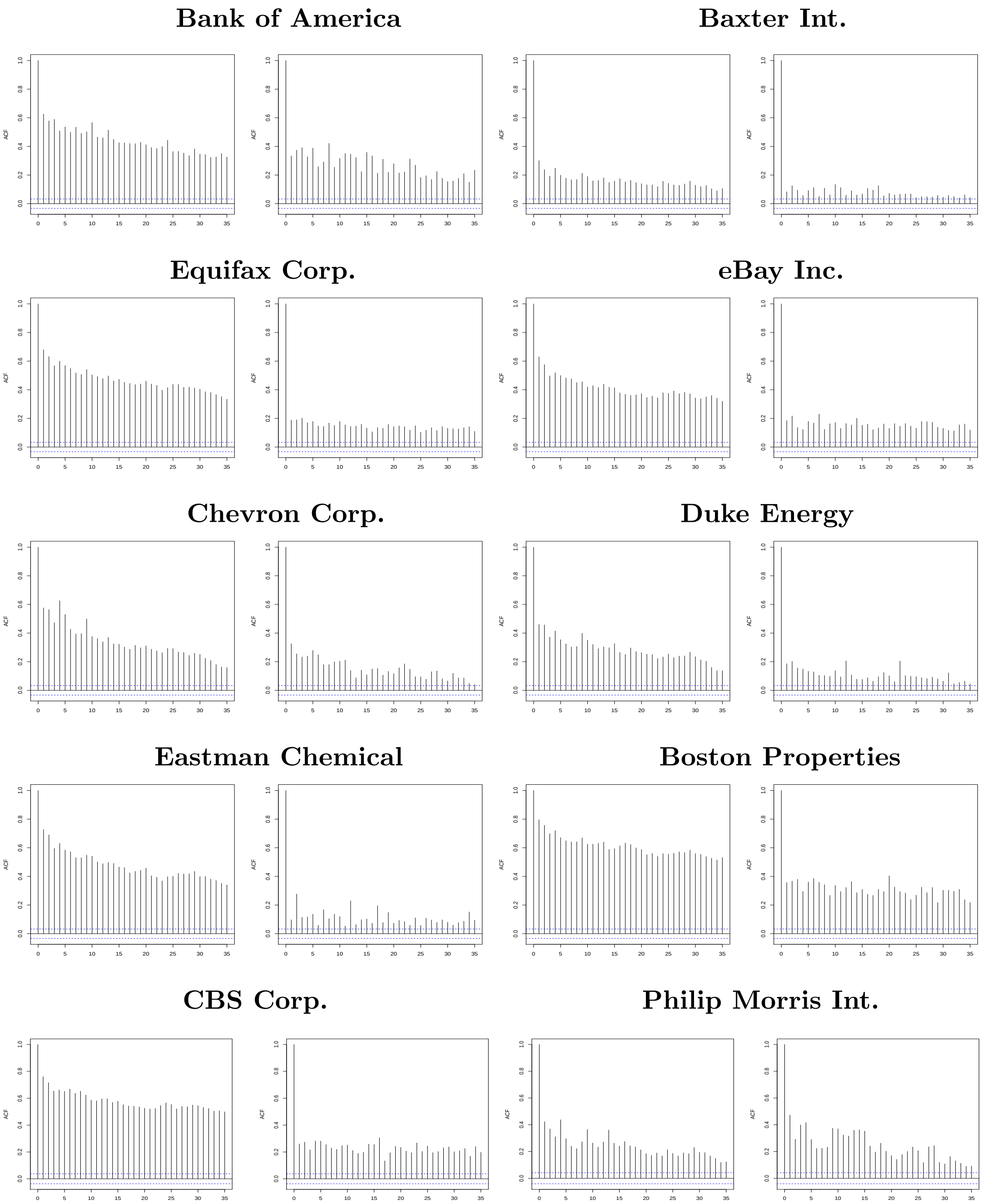
Philip Morris International



(a) RV (b) Jump (c) Intensity

Note: The figure provides the plots for Realized Volatility, Jumps and Intensity of Jumps for representative assets for each Sector. The RV series are computed according Equation (1.3); the jump series instead are computed using Equation (1.18), while the jump intensities according Equation (1.22), where we assumed the explanatory variable is BPV.

Figure 1.3: Realized Volatility, Jump, and Intensity by Sector (cont'd)



(a) RV (b) Jump (c) RV (d) Jump

Note: The Figure contains the autocorrelation functions for Realized Volatility, Jumps for representative assets for each Sector. The jump series are computed using Equation (1.18), while for the computation of the jump intensity, Equation (1.22), we assumed $X_t = BPV_t$.

FIGURE 1.4: Autocorrelation function by Sector

TABLE 1.3: Descriptive statistics of volatility measures

Stock	**Return**	**RV**	**Jumps**	λ	**Stock**	**Return**	**RV**	**Jumps**	λ
JPM	0.0537	0.0363	0.0027	0.5720	BAX	0.0357	0.0155	0.0013	0.5969
	(2.0124)	(0.1076)	(0.0113)	(0.0153)		(1.2008)	(0.0336)	(0.0056)	(0.0782)
BLK	0.0662	0.0310	0.0032	0.6610	ABT	0.0323	0.0146	0.0011	0.6207
	(1.7916)	(0.0682)	(0.0118)	(0.1534)		(1.0818)	(0.0262)	(0.0052)	(0.0542)
BAC	0.0336	0.0521	0.0046	0.6161	JNJ	0.0264	0.0089	0.0005	0.3826
	(2.5993)	(0.1839)	(0.0222)	(0.0000)		(0.8402)	(0.0217)	(0.0041)	(0.0821)
AXP	0.0387	0.0315	0.0024	0.4924	MDT	0.0225	0.0145	0.0011	0.5242
	(1.8326)	(0.0939)	(0.0130)	(0.0491)		(1.2159)	(0.0370)	(0.0110)	(0.0208)
WFC	0.0476	0.0395	0.0031	0.4895	PFE	0.0197	0.0175	0.0015	0.6519
	(2.1363)	(0.1224)	(0.0146)	(0.0780)		(1.1988)	(0.0289)	(0.0048)	(0.0358)
Financial	0.0480	0.0381	0.0032	0.5662	**Health**	0.0273	0.0142	0.0011	0.5553
	(2.07444)	(0.1152)	(0.01458)	(0.05916)		(1.1075)	(0.02948)	(0.00614)	(0.05422)
ALK	0.0841	0.0599	0.0072	0.7291	AAPL	0.1131	0.0379	0.0078	0.6537
	(2.3311)	(0.1027)	(0.0182)	(0.0474)		(1.8264)	(0.0729)	(0.0184)	(0.1316)
EFX	0.0448	0.0171	0.0013	0.5762	EBAY	0.0465	0.0304	0.0028	0.6773
	(1.2894)	(0.0273)	(0.0034)	(0.0681)		(1.8138)	(0.0416)	(0.0063)	(0.0760)
FDX	0.0371	0.0200	0.0012	0.5273	AMZN	0.0961	0.0380	0.0027	0.5882
	(1.4746)	(0.0317)	(0.0038)	(0.0166)		(2.1394)	(0.0623)	(0.0064)	(0.0019)
UNP	0.0543	0.0239	0.0018	0.5971	INTC	0.0344	0.0245	0.0020	0.6616
	(1.4538)	(0.0462)	(0.0057)	(0.0826)		(1.5479)	(0.0387)	(0.0044)	(0.0795)
KSU	0.0586	0.0413	0.0049	0.6841	ADBE	0.0592	0.0276	0.0021	0.6024
	(1.9509)	(0.0730)	(0.0345)	(0.0511)		(1.7720)	(0.0423)	(0.0053)	(0.0004)
Industrial	0.0558	0.0324	0.0033	0.6228	**IT**	0.0699	0.0317	0.0035	0.6366
	(1.69996)	(0.05618)	(0.01312)	(0.05316)		(1.8199)	(0.05156)	(0.00816)	(0.05788)
XOM	0.0311	0.0165	0.0009	0.6610	DUK	0.0354	0.0146	0.0013	0.5805
	(1.2463)	(0.0482)	(0.0043)	(0.1534)		(1.0561)	(0.0317)	(0.0061)	(0.1303)
PXD	0.0643	0.0487	0.0039	0.6161	AEP	0.0346	0.0160	0.0013	0.5597
	(2.1993)	(0.0820)	(0.0179)	(0.0000)		(1.0930)	(0.0360)	(0.0078)	(0.0657)
CVX	0.0418	0.0192	0.0011	0.4924	PPL	0.0322	0.0167	0.0014	0.6202
	(1.3440)	(0.0500)	(0.0041)	(0.0491)		(1.1543)	(0.0381)	(0.0061)	(0.0620)
APA	0.0347	0.0369	0.0025	0.5657	FE	0.0189	0.0177	0.0012	0.5845
	(1.9323)	(0.0657)	(0.0074)	(0.0029)		(1.2450)	(0.0430)	(0.0049)	(0.0010)
SLB	0.0464	0.0351	0.0023	0.5720	DE	0.0273	0.0112	0.0007	0.5202
	(1.8073)	(0.0617)	(0.0077)	(0.0153)		(0.8647)	(0.0218)	(0.0031)	(0.0589)
Energy	0.0437	0.0313	0.0021	0.5815	**Utilities**	0.0297	0.0152	0.0012	0.5730
	(1.70584)	(0.06152)	(0.00828)	(0.04414)		(1.08262)	(0.03412)	(0.0056)	(0.06358)
VMC	0.0441	0.0380	0.0039	0.6157	BXP	0.0543	0.0298	0.0022	0.5426
	(1.8411)	(0.0992)	(0.0618)	(0.0000)		(1.8371)	(0.0739)	(0.0087)	(0.0923)
WRK	0.0408	0.0186	0.0013	0.4895	FRT	0.0541	0.0283	0.0027	0.5978
	(1.3014)	(0.0409)	(0.0058)	(0.0780)		(1.6563)	(0.0734)	(0.0151)	(0.0632)
AVY	0.0244	0.0203	0.0017	0.5812	VTR	0.0639	0.0334	0.0037	0.6240
	(1.4730)	(0.0319)	(0.0076)	(0.0226)		(1.8601)	(0.0761)	(0.0140)	(0.1080)
APD	0.0390	0.0192	0.0012	0.5325	PSA	0.0610	0.0277	0.0030	0.5547
	(1.3578)	(0.0398)	(0.0041)	(0.0451)		(1.7119)	(0.0688)	(0.0256)	(0.1118)
EMN	0.0506	0.0256	0.0022	0.6445	SPG	0.0625	0.0331	0.0030	0.5254
	(1.6533)	(0.0411)	(0.0071)	(0.0462)		(1.9386)	(0.0829)	(0.0178)	(0.0697)
Materials	0.0398	0.0243	0.0021	0.5727	**Real Estate**	0.0592	0.0305	0.0029	0.5689
	(1.52532)	(0.05058)	(0.01728)	(0.03838)		(1.8008)	(0.07502)	(0.01624)	(0.089)
RL	0.0442	0.0314	0.0030	0.6349	PM	0.0390	0.0168	0.0012	0.4658
	(1.8319)	(0.0525)	(0.0111)	(0.0374)		(1.1791)	(0.0641)	(0.0060)	(0.0734)
DIS	0.0510	0.0192	0.0013	0.6519	CVS	0.0490	0.0210	0.0017	0.5451
	(1.4031)	(0.0399)	(0.0045)	(0.0358)		(1.3631)	(0.0652)	(0.0115)	(0.1019)
TSCO	0.0711	0.0387	0.0048	0.6917	K	0.0263	0.0094	0.0005	0.4745
	(1.8225)	(0.0498)	(0.0127)	(0.1296)		(0.9081)	(0.0166)	(0.0023)	(0.0878)
RCL	0.0662	0.0453	0.0037	0.6405	CL	0.0306	0.0102	0.0007	0.5126
	(2.3954)	(0.0901)	(0.0110)	(0.0192)		(0.9785)	(0.0191)	(0.0039)	(0.1035)
CBS	0.0530	0.0451	0.0040	0.6211	KO	0.0242	0.0104	0.0007	0.5691
	(2.1875)	(0.0892)	(0.0144)	(0.0197)		(0.9350)	(0.0201)	(0.0028)	(0.1174)
Cons. Discret.	0.0571	0.0359	0.0034	0.6480	**Cons. Staples**	0.0338	0.0136	0.0010	0.5134
	(0.0571)	(0.03594)	(0.00336)	(0.64802)		(0.03382)	(0.01356)	(0.00096)	(0.51342)

Note: The Table provides the main descriptive statistics for all the assets under analysis. The stocks are divided into 8 Sectors according the Global industry Classification Standard (GICS). In parenthesis are displayed the standard deviation for each of the series. The jump series are computed using Equation (1.18), while for the computation of the jump intensity, Equation (1.22), we assumed $X_t = BPV_t$.

All the volatility series show, with different degree, the widely documented volatility clustering effect. In spite of the idiosyncratic heterogeneity of the data, it is possible to identify similar periods of time in which the volatility has marked significant high level, such as Global Financial Crisis (GFC) and Sovereign Government Bond crisis or low level, e.g. in 2013, where we observe a remarkable increase of the general health of the market.

Consistent with the volatility results, the jump series depicted in the middle panels and computed according to Equation (1.18), exhibit mostly small and positive values of observations. The use of Equations (1.18) and (1.19) helps us to identify only those jumps that significantly affect the volatility and to neglect those that can be considered as measurement errors. Such series also contain a considerable number of extremes values that represent relevant events and market reactions during those days. Particularly affected by jumps are the assets belonging to Industrial, Consumer Discretionary, IT and, as expected, Financial sectors where we observe the highest value of jumps across all the assets, 0.5147. Conversely, the less affected ones appear to be the assets from Health, Utility and Consumer Staples sectors. It is also clear that not only volatility but also jump is a persistent process, alternating calm and wild periods; especially during the GFC. Further alignments are observable with the plots depicting the jump intensity series, λ[13], computed in accordance with Equation (1.22) and reported in the last column. As expected, the series reached their peaks for most of the assets during the financial crisis where the probability of observing a jump at time t given its occurrence at time $t-1$ was around 0.63-0.70. After a long period of financial recovery and a more prosperous economics situation, where the volatility measure shows a significant decrease in term of value, λ reached its minimum in late 2013, when the S&P500 increased around 27% on annual basis and the NASDAQs increased around 35%.

The aforementioned results are confirmed by the summary statistics in Table 1.3, where we report the mean and the standard deviation (in parenthesis) of asset return, Realized Volatility, jump and jump intensity[14]. The analysis of Table 1.3 indicates some key results leading to a better understanding of the main evidence of this paper. Firstly, we note that those sectors that have a higher level of volatility also show a higher level

[13]We report here the results of jump intensity, λ, considering in Equation (1.22) $X_t = BPV_t$. As robustness check we perform all our analysis also with different proxies as suggested by Clements and Liao (2013). Those results are available upon request.

[14] An analytical examination of the descriptive statistics can be found in Appendix B.

of jump components, such as Financial, IT, Consumer Discretionary and Industrial. As expected the two jump components are strictly related with an average correlation of 0.92, suggesting that higher probability of occurrence corresponds to higher jump, in terms of magnitude, and vice versa: e.g. Financial and Consumer Staples, respectively. The results are in line with Bollerslev *et al.* (2009) and show that both volatility and jump measures are highly positively skewed and leptokurtic as shown in Appendix B. Finally, starting from the observation of the presence of clustering effect and serial dependence in both volatility and jump components, we investigate their sample autocorrelation functions, see Figure 1.4. Here we report the autocorrelation function of representative assets, one for each GICS sector. The ACF for the volatility measures across of the sectors exhibits a very slow hyperbolic decay rate with significant coefficients up to 150^{th} lag, clearly suggesting the presence of long memory effect. Conversely, the jumps exhibit a relatively smaller degree of persistence due to the clustering effect, with an average coefficient around 0.23 and most dependency attributable to the first and the fifth lag, corresponding to jumps that are one day and one week apart, respectively. The Hurst long memory parameters confirm these findings of long memory, showing the following values: 0.84 and 0.57 on average, respectively for RV and J across all the assets.

1.6 Empirical analysis

1.6.1 Structural break and long memory

Following Mikosch and Starica (2004), the spurious persistence of volatility indexes, shown in Figure 1.5, might indicate the presence of structural breaks. For such reason, before describing the results of our forecasting exercise, it is convenient to perform a study for the detection of instabilities along the volatility series of each asset. The evidence of occasional breaks in macroeconomic and financial time series, is well documented and underlines the key role played by the structural instability forecasting time-series (see Pesaran and Timmermann (2004) and Pesaran *et al.* (2006)). Ignoring the presence of structural breaks might affect the value of the variables and cause misleading results. Therefore this analysis aims to see whether the relationship between

long memory and structural breaks is present in our data, thus justifying the employment of the long memory models.
To investigate the possibility of having structural breaks in our volatility series, we utilise the pure multiple break in the mean method, from Bai and Perron (2003).[15] Here, the observed process is defined with $M+1$ regimes as

$$z_t = m_j + \varepsilon_t, \qquad t = T_{j-1}+1, T_{j-1}+2, \dots, T_j, \tag{1.26}$$

where $j = 1, 2, \dots, M+1$, z_t is the $\log RV$, m_j is its mean and ε_t the error term, allowed to be serially correlated. In line with financial literature, we decide to assume 8 as the maximum number of breaks in our sample, (30/01/2002 - 31/05/2017) employing the following procedure: firstly, we utilise the $UDmax$ and $WDmax$ statistics to investigate the presence of at least one break, then under evidence of at least one break, we run the sequential $supF_T(l+1|l)$ test[16] to determine the number of structural breaks. The aforementioned procedure is also repeated using the Bayesian Information Criterion (BIC). Table 1.4 displays the test results for the logarithmic transformation of the volatility measures for all assets, while Figure 1.5 reports these series with the respective breaks identified by the procedure.
All the $UDmax$ and $WDmax$ statistics, (not reported in Table 1.4 but available upon request to the author) clearly exhibit the presence of at least one break in all the assets and show the need to move forward performing the $\sup F_T(l+1|l)$ test. The testing results show that all assets have been affected by 4 and 5 breaks along the sample with the exception of CBS Corp. and Coca Cola Ind. where only 3 breaks are identified.

[15] Although, we are aware of the multitude of approaches available to investigate structural breaks (Mikosch and Starica (2004), Hillebrand (2005), Pesaran and Timmermann (2004), Pesaran *et al.* (2006), among the others), we decided to adopt Bai and Perron method to compare the forecasting results with the ones presented by Yang and Chen (2014)

[16] The method, named $\sup F_T(l+1|l)$ consists in the application of $(l+1)$ sup-*Wald* tests for the absence of breaks versus the alternative of a single break. It is used repetitively in each segment containing the observations $\hat{T}_{i-1}$ to $\hat{T}_i$, where $i = 1, \dots, l+1$ to investigate the presence of an additional break. The null hypothesis will be rejected in favour of a model with $l+1$ breaks when "the overall minimal value of the sum of squared residuals (over all segments where an additional break is included) is sufficiently smaller than the sum of squares residuals from the l break model' (Bai and Perron, 2003). The critical values of these test are reported in Bai and Perron (1998).

TABLE 1.4: Multiple structural change tests for log RV

Stock	$supF_T(l+1\mid l)$	**Breaks-BIC**	**Date**	**Stock**	$supF_T(l+1\mid l)$	**Breaks-BIC**	**Date**	**Stock**	$supF_T(l+1\mid l)$	**Breaks-BIC**	**Date**	**Stock**	$supF_T(l+1\mid l)$	**Breaks-BIC**	**Date**	**Stock**	$supF_T(l+1\mid l)$	**Breaks-BIC**	**Date**
JPM	4	4	10/08/2005	XOM	5	5	15/08/2005	RL	5	5	10/08/2005	AAPL	5	4	14/08/2005	BXP	5	5	08/06/2005
			17/07/2007				20/06/2007				24/07/2007				06/07/2007				24/07/2007
			27/08/2009				03/08/2009				03/09/2009				18/08/2009				06/11/2009
			03/08/2012				19/12/2011				16/11/2012				05/06/2012				22/12/2011
			-				24/09/2014				05/01/2015				18/12/2014				30/12/2014
BLK	4	4	08/07/2005	PXD	5	4	23/08/2005	DIS	4	3	10/07/2005	EBAY	5	5	10/08/2005	FRT	5	4	21/08/2005
			16/07/2007				20/05/2008				16/10/2007				02/10/2007				27/06/2007
			26/08/2009				02/07/2010				27/11/2009				12/11/2009				09/11/2009
			27/01/2012				14/08/2012				09/02/2012				27/12/2011				22/12/2011
			-				30/09/2014				-				30/04/2014				11/12/2014
BAC	4	4	16/10/2007	CVX	5	4	16/08/2005	TSCO	4	4	15/08/2005	AMZN	5	5	14/08/2005	VTR	5	4	10/08/2005
			27/11/2009				19/06/2007				23/07/2007				06/07/2007				19/07/2007
			20/11/2012				31/07/2009				01/09/2009				18/08/2009				06/11/2009
			06/01/2015				02/08/2012				03/08/2012				05/06/2012				22/12/2011
			-				26/09/2014				-				18/12/2014				11/12/2014
AXP	4	5	19/05/2005	APA	5	5	15/08/2005	RCL	5	5	10/08/2005	INTC	4	4	11/09/2005	PSA	4	4	17/07/2007
			23/07/2007				23/07/2007				18/10/2007				10/10/2007				07/10/2009
			02/09/2009				02/09/2009				01/12/2009				20/11/2009				19/12/2011
			22/12/2011				03/08/2012				06/08/2012				20/01/2012				30/12/2014
			-				29/09/2014				30/09/2014				-				-
WFC	5	5	10/08/2005	EMN	5	5	24/07/2005	CBS	3	3	23/08/2008	ADBE	4	4	10/08/2005	SPG	5	5	17/07/2005
			23/07/2007				09/10/2007				23/02/2010				24/07/2007				21/06/2007
			02/09/2009				19/11/2009				22/12/2011				02/09/2009				06/11/2009
			21/06/2012				02/08/2012				-				21/12/2011				21/12/2011
			31/12/2014				23/09/2014				-				-				05/01/2015
ALK	4	5	05/08/2005	VMC	4	3	03/07/2005	BAX	5	5	05/08/2005	DUK	5	5	10/07/2005	PM	4	4	21/08/2005
			25/07/2007				19/07/2007				17/07/2007				23/05/2007				27/06/2007
			06/11/2009				06/11/2009				27/08/2009				30/07/2009				09/11/2009
			27/12/2011				03/09/2013				22/12/2011				13/12/2011				22/12/2011
			-				-				31/12/2014				29/09/2014				11/12/2014
EFX	5	5	10/08/2005	WRK	5	5	24/08/2005	ABT	5	5	11/08/2005	AEP	5	4	04/08/2005	CVS	5	4	11/09/2005
			09/10/2007				23/07/2007				20/06/2007				19/06/2007				23/07/2007
			19/11/2009				02/09/2009				03/08/2009				31/07/2009				02/09/2009
			13/08/2012				22/12/2011				09/12/2011				10/11/2011				09/02/2012
			30/12/2014				02/01/2015				01/12/2014				24/07/2014				08/12/2014
FDX	4	4	10/10/2007	AVY	4	4	18/08/2005	JNJ	5	4	11/08/2005	PPL	4	4	04/08/2007	K	5	5	11/08/2005
			20/11/2009				23/07/2007				05/07/2007				05/08/2009				20/06/2007
			23/01/2012				14/09/2009				17/08/2009				10/11/2011				03/08/2009
			26/03/2014				20/11/2012				16/12/2011				05/09/2014				28/11/2011
			-				-				23/09/2014				-				29/09/2014
UNP	5	5	18/07/2005	APD	5	5	24/07/2005	MDT	4	4	07/08/2005	FE	5	5	21/08/2005	CL	4	4	24/06/2005
			24/07/2007				23/07/2007				23/07/2007				06/06/2007				17/07/2007
			03/09/2009				02/09/2009				03/09/2009				03/08/2009				27/08/2009
			03/08/2012				22/12/2011				01/02/2012				21/12/2011				31/01/2012
			30/09/2014				15/09/2014				-				29/09/2014				-
KSU	5	5	18/07/2005	EMN	5	4	01/08/2005	PFE	5	4	22/08/2005	DE	4	3	31/08/2007	KO	4	3	28/06/2005
			25/07/2007				19/07/2007				05/05/2005				14/07/2009				09/07/2007
			03/11/2009				13/11/2009				25/07/2007				30/11/2011				19/08/2009
			08/08/2012				07/08/2012				06/11/2009				07/10/2014				14/12/2011
			29/09/2014				07/10/2014				27/12/2011				-				-

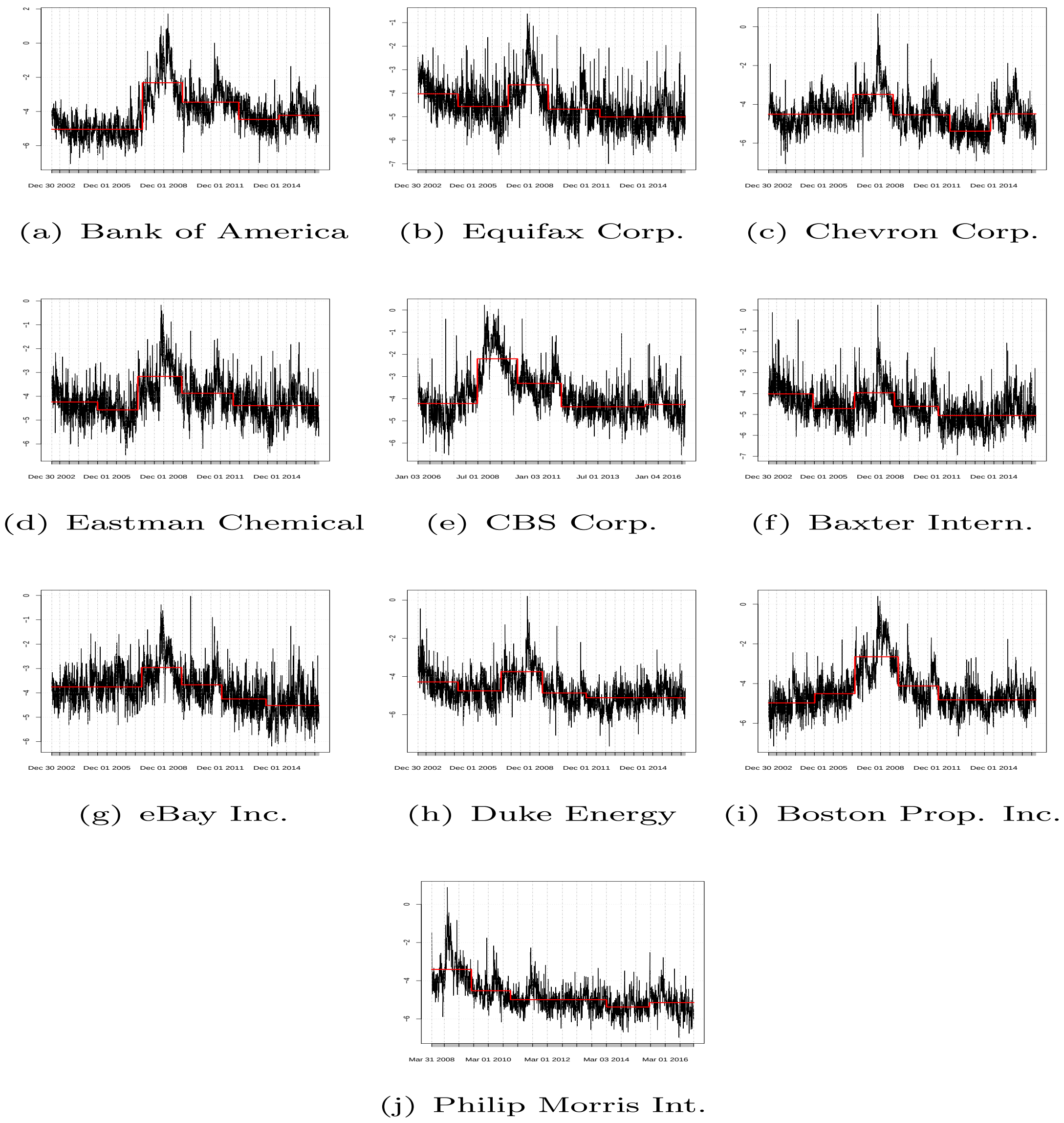
Note: The table provides the results for the structural break tests performed on the log RV of different assets. The first column reports the number of breaks identified by the $supF_T(l+1|l)$ test, Bai and Perron (2003, 2006) at 5% level of significance and discussed in Section 1.6.1. The second column shows instead the number of breaks utilising the minimisation of the Bayesian Information Criterion. In the last column instead there are the break-dates.

TABLE 1.5: Estimation for long memory of $\log RV$

	Before Adj.	**After Adj.**		**Before Adj.**	**After Adj.**
Financial			**Health**		
Average	0.8781	0.8767	Average	0.8367	0.8330
St. dev.	0.0180	0.0187	St. dev.	0.0150	0.0153
Industrial			**IT**		
Average	0.8434	0.8352	Average	0.8505	0.8452
St. dev.	0.0064	0.0099	St. dev.	0.0125	0.0111
Energy			**Utilities**		
Average	0.8785	0.8744	Average	0.85341	0.8497
St. dev.	0.0119	0.0142	St. dev.	0.0153	0.0155
Materials			**Real Estate**		
Average	0.8505	0.8433	Average	0.86502	0.86458
St. dev.	0.0159	0.0167	St. dev.	0.0095	0.0131
Cons. Discr.			**Cons. Stap.**		
Average	0.8499	0.8454	Average	0.83642	0.8322
St. dev.	0.0222	0.0255	St. dev.	0.0069	0.0052

Note: The table provides the averages and the standard deviation of the long memory parameters, calculated by the Local Whittle method, for pre and post the break adjustment for all the Sectors. The table shows the reduction in the long memory parameter due to the adjustments made on the $\log RV$ in order to take into account the breaks identified in Table 1.4.

The analysis of the break-dates suggest that, despite a few misalignments due to firm or sector specific characteristics, all the assets show contemporaneous breaks, mostly attributable to macroeconomic and financial circumstances, as suggested by Beltratti and Morana (2006). Specifically, the first breaks occurred in the spring of 2005. The period was characterised by several events, such as the Hurricane Katrina that had a violent impact on the American economy and the new tax package policy operated by President Bush. In turn, the second and the third dates are related to the Financial crisis of 2007-2009. Indeed, the starting point of the crisis, which manifested as a liquidity crisis, can be dated around the summer 2007, while the the recovery begun in the summer/fall of 2009 in accordance with the NBER dates of recession. The fourth period is mainly related with the European sovereign debt crisis and with the



(a) Bank of America (b) Equifax Corp. (c) Chevron Corp.

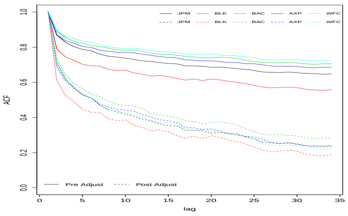
(d) Eastman Chemical (e) CBS Corp. (f) Baxter Intern.

(g) eBay Inc. (h) Duke Energy (i) Boston Prop. Inc.

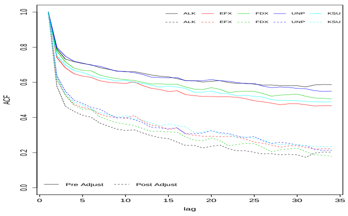
(j) Philip Morris Int.

Note: The figure reports the $\log RV$ plots with the average values of $\log RV$ for every regimes identified using the Bai and Perron test (1998, 2003) with a 5% level of significance, produces for representative assets for each Sector. The use of $\log RV$ allows a clearer representation of the structural breaks; identical results are available for the series of Realized Volatility.

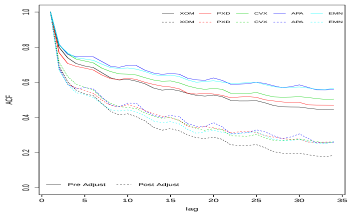
FIGURE 1.5: Structural breaks of Log Realized Volatility by sector



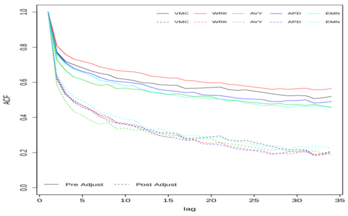
(a) Financial



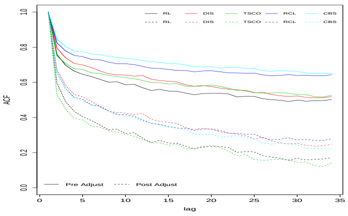
(b) Industrials



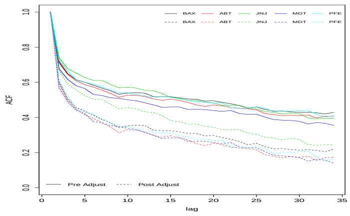
(c) Energy



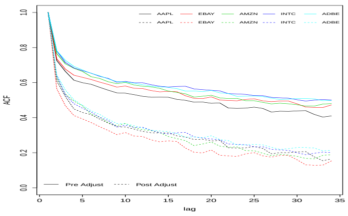
(d) Materials



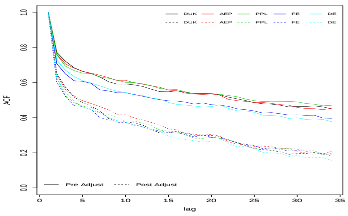
(e) Consumer Discr.



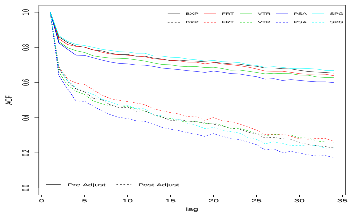
(f) Health



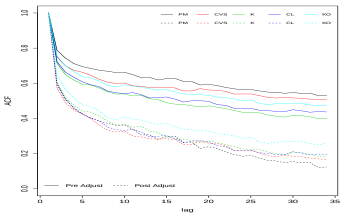
(g) IT



(h) Energy



(i) Real Estate



(j) Consumer Staples

Note: The figure provides the autocorrelation function for the log RV of each asset group by GICS sector, before and after the adjustment made for incorporating the structural breaks as described in Section 1.6.1 We used the Bai and Perron (1998, 2003) test for the identification of structural breaks.

FIGURE 1.6: Comparison of ACF after structural break adjustment

downgrading by Standard and Poor's of the United States sovereign debt from AAA, or "risk free", to AA+, August 2011. The final break point corresponds to a moment in 2013, when the stock market achieved returns on average around 28%, allowing to be considered the best year for stocks since 1997, (S&P 500 has gained 26.8%, Dow Jones Industrial Average has also been impressive with 25.3% return, but the NASDAQs returns have been the most impressive with a rise of 34.5%).
As we expected the volatility series are highly affected by structural breaks that might influence or cause the persistence of the process observed in Figure 1.5. To investigate such possibility, we employ the residual series, $z_t - \hat{m_t}$, Equation (1.26), for incorporating and understanding the effects of structural breaks on the ACF of the assets. Particularly, we quantify their impact observing the variation among the long memory parameters, calculated by the Local Whittle method (LW) before and after the break adjustment. Table 1.5 reports the sector LW averages estimates with the respective standard deviations. We also provide a graphic representation of this effect in Figure 1.6. Both the table and the figure clearly show a remarkable reduction of the serial correlation thanks to break-adjustments, especially for those assets more affected by breaks. Nevertheless, the level of the ACFs remains still high, suggesting that the presence of structural breaks can only partly explain the long memory effect and implicitly confirming the need to use models for capturing such persistence.

1.6.2 Model estimation

The parameter estimates along with the corresponding standard errors for each of the competing models defined in Section 1.4.1 and Appendix A, are reported in Tables 1.6 to 1.8[17]. The models are estimated according the following estimation procedure: ordinary least square estimates with Newey-West covariance correction is used for all methods, except the HAR-FIGARCH, that are estimated using quasi-maximum likelihood estimator. Further, in terms of optimal lag order, we find that a FIGARCH (1,d_u,1) model minimises both the AIC and SIC information criterion for different lag structure combinations; its values are respectively: 3.039 and 3.071. Given the high volume of results and high commonalities among the estimates produced by the models

[17]For the remaining parts of this work, all the volatility measures are computed taking into consideration the correction for the structural breaks, discussed in Section 1.6.1. Results for unadjusted series are available upon request from the author.

for different assets, we decided to present here only three representative assets, that well represents three different degrees of importance of the jump component.
Starting with the results in the first column, the estimates behave in accordance with the literature of HARRV model (see Patton and Shepard (2015)), where the daily, weekly and monthly coefficients are highly statistically significant, with a predominance of the daily lag. Further, α_1, α_2 and α_3 are close to 1, remarking the persistence of the process. Extending the analysis to the volatility components in the other models we have to take into account that we no longer consider RV, but employ Barndorff-Nielsen and Shepard (2007) approach: we focus the attention on the continuous component of volatility, C, defined in Equation (1.19) separated by the jump component. In this new environment, we observe that the contribution of the daily volatility estimates, C decreases significantly once we consider also the jump component; the weekly and monthly volatility components remain unchanged.
Focusing on the jump components, for most of the assets we observe significant parameter estimates associated only with the daily and weekly lags, in accordance with the Bollerslev *et al.* (2009). Such evidence becomes recursively clear when we consider assets more affected by jumps, such as the ones in the Financial, IT, Consumer Discretionary and Industrial sectors. The importance of the jumps is also underlined by the relevance of their intensity, λ, as shown in Equation (1.22). The intensity estimates are positive and statistically significant, despite small in magnitude. This implies that when jumps are more likely to occur, the volatility is higher. The small magnitude of the estimates is a natural consequence of the definition of the intensity of jumps proposed by Clements and Liao (2013). All the jump estimates show a uniformly significant negative sign, confirming the results of Patton and Shepard (2015). This reveals that days dominated by negative jumps lead to higher future volatility, while days with positive jumps lead to lower future volatility. The negative jump component is larger in terms of magnitude of jumps than the positive one for the 2 models proposed by Patton and Shepard (2015), indicating that the impact on future volatility will be higher in magnitude following a negative jump rather than positive one[18].
Further, the asymmetry parameter estimates reveal the strong significance of the negative returns at almost all lags, confirming an heterogeneous structure of the leverage effect. As expected, the negative sign indicates that a lagged negative return shock

[18]Such findings confirms indirectly our decision to consider a restricted model, with only the negative component, rather than both.

leads to an increase in the volatility larger than the one produced by a positive shock of the same magnitude. Further, the magnitude of the estimates supports the view of leverage effect as a key feature for volatility description and an explanation for the change in the sign of the constant term, that becomes negative. The latter results mirror the findings presented in Martens *et al.* (2004) and Corsi *et al.* (2012).
Finally, in accordance with Patton and Shepard (2015), the coefficient of negative semivariance SV has a larger and more significant impact on future volatility than of positive semivariance for all the assets. In fact, the coefficient, SV^+, is not significantly different from zero for weekly and monthly horizons for the 85% of our assets and its magnitude is always lower than the one showed by negative SV and it is particularly evident for those assets more affected by jumps.

1.6.3 Forecasts and evaluation

The availability of a number of volatility models raised the issue of evaluating them accordingly to their forecasting performances to identify the best methodology. For such purpose, we decide to employ two different predictive ability tests: the Model Confidence Set (MCS) procedure recently developed by Hansen *et al.* (2011) and the Giacomini and White (2006) (GW)[19] test. The former procedure consists of a sequence of tests which permit to construct a set of superior models, where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a certain confidence level. The EPA test statistic test is calculated for an arbitrary loss function, implying that we could test models on various aspects depending on the loss function chosen. The GW test, instead, helps to identify whether the differences in forecasting performance of competing models are statistically significant. This test is a t-test with robust standard error. In this setting, the null hypothesis is that the two competing forecasting models have the same predictive accuracy, while the alternative is that the second method performs better. We adopt a rolling window out of sample predictive ability approach[20]. Specifically for the latter, we divide the sample of T trading days in set of

[19]Allowing for an unified analysis of both nested and non-nested models, Giacomini and White (2006), GW, suggested to focus on the core of the test on conditional expectation of forecast errors rather than unconditional, as recommended by Diebold and Mariano (1995). Giacomini and White observed that when the estimation sample is fixed, (parameters are estimated using rolling window data), their test remains asymptotically valid even for nested model.

[20]In Appendix C we display the main results for an in sample forecasting exercise.

H in-sample observations from January, 30, 2002 to December, 31, 2015 and $K = T - H$ out-of-sample observations from January, 02, 2015 to May, 30, 2017. Hence, a rolling window of H observations, $[1 + t : (H + t)]_{t=0}^{K}$ is used to re-estimate the models and produce $K - 1$ out-of-sample day-ahead forecasts. The forecasting performances are compared by analysing the deviation between the volatility forecasts and the actual market volatility proxy, RV. In order to test the robustness of our forecasting results, we consider two different loss functions: Root-Mean Square Error (RMSE) and Gaussian Quasi likelihood (QLIKE)[21].

The analysis of the forecasting performances starts with the reporting of the loss function results for each asset, presented in Table 1.9. Given the high number of tables and figures produced during such analysis, we decided to report in the main body of the paper only the results for RMSE loss function. The findings for QLIKE loss function are reported in Appendix D. The analysis of Table 9 helps us to identify some preliminary findings. Firstly, it is clear that the LHARCJI is the most appropriate volatility model for most of the assets. It produces a reduction in all the loss functions of around 5%-8% on average across the assets with peak of 10% for those assets which are most affected by jumps such as the Industrials and Financials. It appears also being adequate for assets less affected by jumps, such as Kellogg Co.. Surprisingly, the basic HARRV model is not always outperformed by the more sophisticated ones in terms of loss function reduction. Such evidence is also observed for those assets that are less affected by jumps, both in terms of magnitude and intensity, such as Abbott Laboratories and Philip Morris Internationals. In accordance with Clements and Liao (2013), we observe remarkable reduction in the loss functions for those assets heavily affected by jumps when we incorporate the jump intensity. Nevertheless, it is not possible to assert that one of the components drags more forecasting importance, as in Clements and Liao (2013). In other words, we can only confirm that the best way to harness the jump component for forecasting the volatility of these two indexes is to make use of

[21] Denote by σ_t the actual value of volatility, and by h_t the predicted one with $t = \tau + 1, \ldots, T$. Based on these quantities we define the loss functions:

$$RMSE = \sqrt{T^{-1} \sum_{t=\tau+1}^{n} (\sigma_t - h_t)^2}; \qquad QLIKE = T^{-1} \sum_{t=\tau+1}^{T} (\log(h_t^2) - \sigma_t^2 h_t^{-2}).$$

The first loss function represent a standard approach for comparing forecasting performances of different models while QLIKE, has been included, since it has been proved to be robust to noise in the proxy for volatility by Patton (2011) and to have useful properties as described by Patton and Sheppard (2009).

both the magnitude and intensity of jumps, since this leads to reduction in all the loss functions that is statistically significant. Another notable finding is the eminent role of the leverage effect in the prediction of volatility. Despite the method used to capture it, all the models that incorporate the asymmetric impact of past returns, exhibit a remarkable reduction of the loss functions with respect to those models that do not consider it. Although its importance, such forecasting analysis exclusively produces a general indication about the relative forecasting accuracy, without implying whether differences in performance among the models are significant. For this reason, we decide to perform a pairwise analysis of the forecasting models' performances, using the GW test of conditional predictive accuracy. The Tables 1.10 and 1.11 exhibit the average *p-values* at sector level; in other words we report the averages across the assets belonging to the same sector. Here, the null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that the method in the row performs better than the other[22]. Further, we report for each sector and each model also the average loss function.

The analysis of the Tables 1.10 and 1.11 confirms the performance results observed above. For what concerns the LHARCJI model, we notice that all the differences among the models RMSE loss functions, shown in Table 1.9, are statistically significant at 5% level of significance. Hence, the null hypothesis of equal accuracy of forecasting performances is always rejected, implying that the benchmark models (the ones in column) are most of the times outperformed by LHARCJI model. Such results clearly confirm the importance of considering all the volatility features simultaneously in modelling financial volatility. The LHARCJI appears to be the best method in 7 of the 10 sectors, with a rate of success of 90%. In the other sectors, less advanced approaches seem to be more suitable, such as HARRV for Health sector, LHARCJ for Consumer Staples and HARCJ for Utility sector. The latter results are not very surprising and can be motivated by the fact that these industries are the ones with least presence of jumps both in terms of magnitude and intensity. Further, we note that the relevance of the intensity component is not always clear and unquestionable. This is evident in sectors such as Materials, Real Estate and Energy, where the presence of jumps is not predominant. Here, the difference between the RMSE loss functions for models with and without the intensity, λ, appear to be not statistically significant (i.e.

[22]Due to the significant amount of tables produced, we do not report the results for asset level. All the tables are available upon request to the authors.

the two approaches appear to have similar predictive power). Conversely, the leverage effect plays an important role in volatility prediction, with the only exception of the Health sector due to its peculiarities. Incorporating the asymmetric impact of past returns in jump models reduces the loss functions and these reductions are statistically significant at the 5% level of significance.

Finally, we employ the MCS selection procedure, developed by Hansen *et al.* (2011) which instead of a pairwise comparison, introduces a method for evaluating multiple hypothesis simultaneously. The method allows to built a set of superior forecasting models without imposing a benchmark model. This is done by performing a sequence of significance tests, where models that are found to be significantly inferior to the others are eliminated[23]. Table 1.12 reports the summary of MCS results where we show for each assets whether the model under analysis is included in the MCS subset for the out of sample exercise and in case in which position[24]. The main results is that the ranking reflects the different degree with the jumps affect the assets: for highly jumpy assets (e.g. IT and Industrial) the MCS procedure is performing well and is able to identify a unique model in more than 50% of the times. Once the jump component decreases in terms of magnitude and occurrences, simultaneously, we observe an increase in the number of selected models and a decrease in the power of the MCS procedure, implying a latent difficulty in the identification of a single best approach.

Medium affected assets show on average a surviving rate of the 40% while the ones weakly affected by jumps of the 75% with peak of the 90%. The results in Table 1.12 confirm that LHARCJI is one of the best performing models since it is included in the MCS subset for almost every single asset, being also sometimes the only approach survived under the selection procedure, such for FedEx Corp. and Royal Caribbean Cruise Ltd. The analysis of the ranks confirms the above results showing that, if the LHARCJI is included in the MCS sub−set such model has a probability to be the best one in terms of loss functions of around 75%. Also, the LHARCJ volatility model seems to be fitting to the data well, being selected in the 62% of the cases and succeeding in the 22% of those. Such result supports the ambiguity of the jump intensity role in the assets which are less affected by jumps. In line with the previous literature, we observe that the HARRV model has a probability of 80% of being selected by the MCS

[23]In accordance with Hansen *et al.* (2011), the confidence level for the MCS is set to $\alpha = 0.2$ while the number of bootstrap iteration for computing the distribution under the null hypothesis is 5000.

[24]As shown by Hansen *et al.* (2011), their procedure can also be adopted in the sample forecasting studies

procedure for those sectors less affected by jumps while this probability decreases drastically when we enlarge the focus to the other sectors where the jumps are a important component of data. It is very interesting to observe that such basic model, HARRV, is able to achieve the first position of the ranking in around 5% of the time when it has been selected. The test also confirms the key role played by the asymmetric behaviour of returns and volatility. This follows from the fact that the survival rates for those models that take into account the leverage effect is almost three times higher than the ones that do not consider it. Table 1.12 also shows that the survival rate of the forecasting models proposed by Patton and Shepard (2015) is above the 50% for both of them. Further, their approaches appear to fit well to those assets that do not have a clear jump patterns, although without outperforming the LHARCJI approach.
Overall, the analysis conducted in this paper indicates that the LHARCJI as the best forecasting volatility model for most of the assets and sectors, as underlined by the different tests. The cases where the LHARCJI appears not to be the best approach, can be easily explained by specific features of the assets, such as low magnitude of jumps or low persistence of the process.

1.7 Concluding remarks

The understanding of volatility phenomenon is one of the most critical issues in the financial literature. Recently, the strand of research in financial volatility has benefited from the widespread use of high frequency price data, that allowed the computation of a model-free measurement of volatility, named Realized Volatility. In sum, exploiting this new measure of volatility, our work contributes to this literature by introducing a new approach for forecasting financial volatility that synthesizes different well-established volatility characteristics to improve the predictive accuracy of the financial volatility.
Specifically, exploiting the asymptotic properties of the Realized Volatility, we propose a linear model for volatility, named Leverage Heterogeneous Autoregressive Continuous, Jump, Intensity model, LHARCJI, that takes into account the persistence of the process, the asymmetric behaviour of returns and volatility, the magnitude, the sign of the jumps and the probability of their occurrences.
Our empirical analysis not only produces an overwhelming evidence of the need to consider simultaneously all the volatility features but also shows an interesting relation

between level of jumps and the model selected. We find that the most predictive power of our approach is displayed in those assets which are more affected by jumps, such as Industrials, IT and Financials, where we observe a reduction in the loss functions of around 5%-8% on average across the assets, with peak of reduction of 10%. Further, the LHARCJI model appears to be the best one for prediction in 7 of 10 sectors.

Despite the empirical nature of our work, we strongly believe that there is space for further development. From a modelling point of view, further studies could focus on the investigation of new characteristics of volatility, especially in those sectors where the jump component is not a key factor. For instance, accounting for the different reactions of volatility to the arrival of news in terms of speed and magnitude of the impact. Finally, it might be also relevant analysing how improvements in the description of volatility affect the measure of systematic risk. Particularly, it would be interesting to see if our new models may be a valid enhancements of the well established GARCH Dynamic conditional estimator in the extension of the Marginal Expected Shortfall made by Brownlees and Engle (2010).

Table 1.6: Estimation results for FedEx Corporation

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ	LHAR-CJI
$\log RV_{t-1}^{d}$	0.3779	0.2799	0.2997				0.2594		
	(0.0442)	(0.0442)	(0.0199)				(0.0445)		
$\log RV_{t-1}^{w}$	0.545	0.3124	0.5422		0.5419	0.5051	0.5755		
	(0.0743)	(0.0410)	(0.0741)		(0.0741)	(0.0448)	(0.0763)		
$\log RV_{t-1}^{m}$	0.4947	0.4618	0.4921		0.3773	0.5058	0.4264		
	(0.0734)	(0.0439)	(0.0732)		(0.0744)	(0.0586)	(0.0860)		
$\log C_{t-1}^{d}$				0.2879				0.2613	0.2608
				(0.0241)				(0.0444)	(0.0444)
$\log C_{t-1}^{w}$				0.4173				0.5729	0.5732
				(0.0325)				(0.0761)	(0.0761)
$\log C_{t-1}^{m}$				0.3314				0.3233	0.3129
				(0.0280)				(0.0858)	(0.0865)
$\log RS_{t-1}^{+}$					-0.0412				
					(0.0172)				
$\log RS_{t-1}^{-}$					0.2791				
					(0.0442)				
$\log BVP_{t-1}^{d}$						0.3252			
						(0.0205)			
$\log(J_{t-1}^{d-}+1)$		0.0582		0.0413			0.0427	0.069	-0.0691
		(0.0310)		(0.0272)			(0.0595)	(0.0279)	(0.0279)
$\log(J_{t-1}^{w-}+1)$		0.2460		0.2301			0.2801	0.2038	-0.2017
		(0.0723)		(0.0662)			(0.1145)	(0.0677)	(0.0677)
$\log(J_{t-1}^{m-}+1)$		0.5670		0.5634			0.5199	0.5885	-0.5917
		(0.1165)		(0.1408)			(0.1935)	(0.1415)	(0.1416)
λ_t							0.0034		0.0054
							(0.0005)		(0.0006)
ΔJ_{t-1}^{2+}						-0.0082			
						(0.0015)			
ΔJ_{t-1}^{2-}						-0.4046			
						(0.0623)			
r_t^{d-}								-0.006	-0.010
								(0.0008)	(0.0008)
r_t^{w-}								0.0048	0.0051
								(0.0023)	(0.0023)
r_t^{m}								-0.0057	-0.0054
								(0.0045)	(0.0045)
z_t^{d}			-0.0048						
			(0.0009)						
z_t^{w}			-0.0052						
			(0.0014)						
z_t^{m}			-0.0064						
			(0.0008)						
$\lvert z_t^{d}\rvert$			0.0143						
			(0.0056)						
$\lvert z_t^{w}\rvert$			0.0083						
			(0.0123)						
$\lvert z_t^{m}\rvert$			0.1016						
			(0.0289)						
d_u			0.22953						
			(0.0357)						
ω			0.6099						
			(0.1840)						
β_1			0.6138						
			(0.0719)						
ϕ_1			0.4505						
			(0.0655)						
Constant	0.0000	0.0001	-0.0007	0.0001	-0.0003	0.0000	-0.0006	0.0000	-0.0002
	(0.0000)	(0.0000)	(0.0003)	(0.0000)	(0.0003)	(0.0000)	(0.0003)	(0.0000)	(0.0003)
R^2	0.6259	0.6597	0.6046	0.6597	0.6599	0.6641	0.6110	0.6541	0.6643
Adj. R^2	0.625	0.36586	0.6039	0.6586	0.6586	0.6625	0.6097	0.6525	0.6625

Note: The Table presents the estimates for different volatility models, see details in Appendix A. The models are estimated according the following estimation procedure: ordinary least square with Newey-West covariance correction for all methods except the HAR FIGARCH, that are estimated using quasi-maximum likelihood estimator. We find that a FIGARCH (1,d_u,1) model minimises both the AIC and SIC information criterion for different lag structure combinations; its values are respectively: 3.039 and 3.071.

Table 1.7: Estimation results for Eastman Chemicals

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ	LHAR-CJI
$\log RV^{d}_{t-1}$	0.2389	0.1431	0.1235				0.1200		
	(0.0200)	(0.0223)	(0.0194)				(0.0219)		
$\log RV^{w}_{t-1}$	0.2906	0.1715	0.1751		0.1680	0.0843	0.1057		
	(0.0319)	(0.0485)	(0.0413)		(0.0476)	(0.0416)	(0.0482)		
$\log RV^{m}_{t-1}$	0.1901	0.4968	0.4757		0.4903	0.1923	0.2367		
	(0.0274)	(0.0633)	(0.0478)		(0.0618)	(0.0628)	(0.0714)		
$\log C^{d}_{t-1}$				0.1399				0.1171	0.1166
				(0.0218)				(0.0215)	(0.0215)
$\log C^{w}_{t-1}$				0.1685				0.1025	0.0996
				(0.0476)				(0.0472)	(0.0472)
$\log C^{m}_{t-1}$				0.491				0.2312	0.2168
				(0.0618)				(0.0702)	(0.0704)
$\log RS^{+}_{t-1}$					-0.0316				
					(0.0121)				
$\log RS^{-}_{t-1}$					0.1227				
					(0.0242)				
$\log BVP^{d}_{t-1}$						0.1018			
						(0.0192)			
$\log(J^{d-}_{t-1}+1)$		0.0382		0.0432			-0.1048	0.0218	-0.0218
		(0.0101)		(0.0521)			(0.0587)	(0.0515)	(0.0515)
$\log(J^{w-}_{t-1}+1)$		0.1227		0.1227			-0.0881	0.1087	-0.1275
		(0.1240)		(0.1240)			(0.1414)	(0.1236)	(0.1235)
$\log(J^{m-}_{t-1}+1)$		0.1908		0.1908			-0.1854	-0.1114	-0.1763
		(0.2231)		(0.2231)			(0.2598)	(0.2322)	(0.2334)
λ_t							0.0089		0.0104
							(0.0011)		(0.0042)
ΔJ^{2+}_{t-1}						-0.0172			
						(0.0048)			
ΔJ^{2-}_{t-1}						-0.3070			
						(0.0084)			
r^{d-}_t								-0.0017	-0.0016
								(0.0018)	(0.0018)
r^{w-}_t								-0.0307	-0.0307
								(0.0044)	(0.0044)
r^{m}_t								-0.031	-0.0336
								(0.0082)	(0.0083)
z^{d}_t			-0.0022						
			(0.0007)						
z^{w}_t			-0.0071						
			(0.0011)						
z^{m}_t			-0.0089						
			(0.0008						
$\|z^{d}_t\|$			0.0157						
			(0.0066)						
$\|z^{w}_t\|$			0.0053						
			(0.0023)						
$\|z^{m}_t\|$			0.087						
			(0.0311)						
d_u			0.2783						
			(0.0460)						
a_1			0.7654						
			(0.1430)						
β_1			0.6138						
			(0.0719)						
ϕ_1			0.4001						
			(0.0308)						
Constant	0.0000	0.0001	−0.0002	0.0001	−0.0002	−0.0001	−0.0008	−0.0001	−0.0008
	(0.0000)	(0.0000)	(0.0003)	(0.0000)	(0.0003)	(0.0000)	(0.0003)	(0.0000)	(0.0003)
R^2	0.4975	0.5631	0.5623	0.5628	0.5631	0.5978	0.5969	0.5974	0.6088
Adj. R^2	0.4972	0.5616	0.5613	0.5613	0.5614	0.5958	0.5953	0.5954	0.6066

Note: The Table presents the estimates for different volatility models, see details in Appendix A. The models are estimated according the following estimation procedure: ordinary least square with Newey-West covariance correction for all methods except the HAR FIGARCH, that are estimated using quasi-maximum likelihood estimator. We find that a FIGARCH (1,d_u,1) model minimises both the AIC and SIC information criterion for different lag structure combinations; its values are respectively: 3.039 and 3.071.

Table 1.8: Estimation results for Baxter International Inc.

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ	LHAR-CJI
$\log RV^{d}_{t-1}$	0.3544	0.1188	0.3338				0.1117		
	(0.0407)	(0.0196)	(0.0192)				(0.0412)		
$\log RV^{w}_{t-1}$	0.3118	0.2175	0.2658		0.2135	0.2030	0.1511		
	(0.0908)	(0.0412)	(0.0453)		(0.0907)	(0.0569)	(0.0950)		
$\log RV^{m}_{t-1}$	0.4218	0.4391	0.3970		0.2920	0.2813	0.1803		
	(0.1132)	(0.0473)	(0.0668)		(0.0946)	(0.1043)	(0.1328)		
$\log C^{d}_{t-1}$				0.1541				0.1114	0.1109
				(0.0407)				(0.0412)	(0.0412)
$\log C^{w}_{t-1}$				0.2120				0.2512	0.2511
				(0.0907)				(0.0949)	(0.0948)
$\log C^{m}_{t-1}$				0.3214				0.2799	0.2242
				(0.1131)				(0.0927)	(0.0954)
$\log RS^{+}_{t-1}$					-0.0761				
					(0.3325)				
$\log RS^{-}_{t-1}$					0.286				
					(0.0469)				
$\log BVP^{d}_{t-1}$						0.3038			
						(0.0218)			
$\log(J^{d-}_{t-1}+1)$		0.0282		0.0233			-0.0303	-0.0184	-0.0181
		(0.0094)		(0.0284)			(0.0573)	(0.0312)	(0.0311)
$\log(J^{w-}_{t-1}+1)$		0.0946		-0.1211			-0.1244	-0.1037	-0.1021
		(0.0223)		(0.0709)			(0.1335)	(0.0806)	(0.0806)
$\log(J^{m-}_{t-1}+1)$		0.1067		0.1186			-0.1657	0.1149	0.0761
		(0.0165)		(0.1259)			(0.1973)	(0.1513)	(0.1525)
λ_t							0.0037		0.00246
							(0.0018)		(0.0083)
ΔJ^{2+}_{t-1}						-0.0580			
						(0.0224)			
ΔJ^{2-}_{t-1}						-0.1197			
						(0.0347)			
r^{d-}_t							-0.0087	-0.0087	-0.0087
							(0.0022)	(0.0022)	(0.0022)
r^{w-}_t							-0.0115	-0.0115	-0.0117
							(0.0054)	-0.0052	-0.0052
r^{m}_t							-0.0033	-0.0033	-0.0058
							(0.0103)	(0.0103)	(0.0103)
z^{d}_t			-0.0062						
			(0.0014)						
z^{w}_t			-0.0094						
			(0.0027)						
z^{m}_t			-0.0076						
			(0.0031)						
$\lvert z^{d}_t \rvert$			0.0267						
			(0.0084)						
$\lvert z^{w}_t \rvert$			0.00249						
			(0.0008)						
$\lvert z^{m}_t \rvert$			0.0604						
			(0.0091)						
d_u			0.2183						
			(0.0689)						
a_1			0.4035						
			(0.0726)						
β_1			0.3149						
			(0.0270)						
ϕ_1			0.2011						
			(0.0097)						
Constant	0.0000	0.0001	−0.0001	0.0001	−0.0001	0.0001	−0.0002	0.0001	−0.0002
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0001)
R^2	0.3658	0.4384	0.4311	0.4384	0.4391	0.4493	0.4408	0.4494	0.4705
Adj. R^2	0.3651	0.4368	0.4300	0.4368	0.4372	0.4469	0.4389	0.4469	0.4678

Note: The Table presents the estimates for different volatility models, see details in Appendix A. The models are estimated according the following estimation procedure: ordinary least square with Newey-West covariance correction for all methods except the HAR FIGARCH, that are estimated using quasi-maximum likelihood estimator. We find that a FIGARCH $(1,d_u,1)$ model minimises both the AIC and SIC information criterion for different lag structure combinations; its values are respectively: 3.039 and 3.071.

Table 1.9: Loss function (RMSE) results for all assets

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ	LHAR-CJI
Financial	0.0239	0.0216	0.0229	0.0237	0.0159	0.0191	0.0119	0.0127	0.0063
JPM	0.0167	0.0259	0.0202	0.0225	0.0148	0.0257	0.0170	0.0179	0.0069
BLK	0.0334	0.0147	0.0288	0.0173	0.0125	0.0141	0.0130	0.0122	0.0059
BAC	0.0381	0.0328	0.0310	0.0389	0.0209	0.0328	0.0112	0.0170	0.0054
AXP	0.0156	0.0141	0.0171	0.0134	0.0108	0.0141	0.0079	0.0068	0.0059
WFC	0.0156	0.0207	0.0176	0.0263	0.0204	0.0089	0.0103	0.0094	0.0072
Industrial	0.0272	0.0170	0.0247	0.0184	0.0172	0.0138	0.0153	0.0141	0.0083
ALK	0.0182	0.0205	0.0188	0.0181	0.0183	0.0156	0.0199	0.0138	0.0097
EFX	0.0113	0.0141	0.0131	0.0145	0.0134	0.0105	0.0111	0.0120	0.0083
FDX	0.0145	0.0082	0.0145	0.0129	0.0145	0.0070	0.0076	0.0120	0.0050
UNP	0.0148	0.0141	0.0150	0.0132	0.0132	0.0116	0.0125	0.0128	0.0092
KSU	0.0772	0.0282	0.0623	0.0333	0.0265	0.0244	0.0254	0.0197	0.0095
Energy	0.0146	0.0167	0.0207	0.0206	0.0168	0.0133	0.0133	0.0125	0.0096
XOM	0.0075	0.0119	0.0115	0.0121	0.0122	0.0075	0.0073	0.0065	0.0051
PXD	0.0236	0.0222	0.0205	0.0292	0.0221	0.0203	0.0203	0.0215	0.0145
CVX	0.0087	0.0108	0.0107	0.0108	0.0108	0.0079	0.0079	0.0067	0.0047
APA	0.0215	0.0277	0.0486	0.0390	0.0281	0.0220	0.0217	0.0189	0.0182
EMN	0.0116	0.0110	0.0120	0.0122	0.0110	0.0087	0.0090	0.0091	0.0054
Material	0.0204	0.0202	0.0222	0.0197	0.0183	0.0144	0.0159	0.0125	0.0114
VMC	0.0254	0.0222	0.0274	0.0199	0.0214	0.0185	0.0194	0.0165	0.0159
WRK	0.0156	0.0153	0.0166	0.0172	0.0150	0.0116	0.0118	0.0118	0.0103
AVY	0.0258	0.0215	0.0266	0.0206	0.0179	0.0147	0.0178	0.0150	0.0139
APD	0.0244	0.0144	0.0175	0.0196	0.0143	0.0116	0.0120	0.0106	0.0094
EMN	0.0106	0.0276	0.0227	0.0210	0.0227	0.0155	0.0184	0.0085	0.0076
Cons. Discr.	0.0563	0.0262	0.0516	0.0298	0.0250	0.0220	0.0229	0.0195	0.0129
RL	0.0676	0.0321	0.0615	0.0376	0.0319	0.0277	0.0279	0.0260	0.0177
DIS	0.0563	0.0262	0.0516	0.0298	0.0250	0.0220	0.0229	0.0195	0.0129
TSCO	0.1099	0.0234	0.0966	0.0312	0.0207	0.0233	0.0251	0.0175	0.0080
RCL	0.0291	0.0313	0.0305	0.0284	0.0301	0.0218	0.0228	0.0174	0.0141
CBS	0.0187	0.0178	0.0177	0.0220	0.0174	0.0152	0.0156	0.0172	0.0120
Health	0.0091	0.0148	0.0191	0.0156	0.0138	0.0100	0.0112	0.0102	0.0092
BAX	0.0064	0.0168	0.0154	0.0159	0.0152	0.0101	0.0119	0.0103	0.0075
ABT	0.0110	0.0155	0.0139	0.0136	0.0146	0.0105	0.0113	0.0096	0.0103
JNJ	0.0029	0.0084	0.0085	0.0106	0.0086	0.0051	0.0052	0.0060	0.0049
MDT	0.0123	0.0195	0.0178	0.0203	0.0185	0.0129	0.0149	0.0121	0.0118
PFE	0.0131	0.0140	0.0400	0.0176	0.0124	0.0115	0.0129	0.0131	0.0113
IT	0.0337	0.0176	0.0320	0.0203	0.0139	0.0142	0.0135	0.0138	0.0090
AAPL	0.0251	0.0160	0.0296	0.0182	0.0140	0.0120	0.0125	0.0128	0.0088
EBAY	0.0791	0.0186	0.0713	0.0248	0.0145	0.0136	0.0144	0.0158	0.0091
AMZN	0.0262	0.0221	0.0241	0.0229	0.0171	0.0174	0.0185	0.0196	0.0106
INTC	0.0171	0.0133	0.0160	0.0129	0.0131	0.0114	0.0089	0.0103	0.0051
ADBE	0.0209	0.0182	0.0192	0.0225	0.0106	0.0166	0.0134	0.0104	0.0113
Utilities	0.0110	0.0116	0.0116	0.0121	0.0104	0.0081	0.0089	0.0075	0.0070
DUK	0.0077	0.0084	0.0082	0.0117	0.0075	0.0049	0.0057	0.0071	0.0041
AEP	0.0104	0.0132	0.0122	0.0125	0.0108	0.0083	0.0096	0.0060	0.0052
PPL	0.0139	0.0101	0.0126	0.0099	0.0086	0.0082	0.0090	0.0064	0.0094
FE	0.0145	0.0188	0.0167	0.0199	0.0179	0.0133	0.0138	0.0129	0.0115
DE	0.0084	0.0077	0.0083	0.0067	0.0072	0.0060	0.0064	0.0052	0.0048
Real estate	0.0095	0.0234	0.0226	0.0222	0.0228	0.0095	0.0102	0.0070	0.0068
BXP	0.0087	0.0221	0.0215	0.0224	0.0221	0.0083	0.0094	0.0065	0.0057
FRT	0.0064	0.0156	0.0140	0.0159	0.0156	0.0061	0.0072	0.0056	0.0047
VTR	0.0143	0.0418	0.0411	0.0385	0.0391	0.0164	0.0162	0.0088	0.0100
PSA	0.0106	0.0189	0.0182	0.0178	0.0187	0.0096	0.0098	0.0078	0.0076
SPG	0.0077	0.0187	0.0180	0.0166	0.0185	0.0069	0.0083	0.0064	0.0061
Cons. Stap.	0.0164	0.0162	0.0181	0.0176	0.0153	0.0093	0.0109	0.0078	0.0081
PM	0.0049	0.0137	0.0132	0.0134	0.0132	0.0072	0.0084	0.0052	0.0058
CVS	0.0476	0.0279	0.0396	0.0379	0.0273	0.0156	0.0196	0.0149	0.0160
K	0.0119	0.0132	0.0134	0.0129	0.0131	0.0111	0.0112	0.0101	0.0109
CL	0.0083	0.0164	0.0160	0.0150	0.0148	0.0085	0.0096	0.0059	0.0051
KO	0.0094	0.0099	0.0081	0.0087	0.0082	0.0040	0.0057	0.0029	0.0026

Note: The Table provides the root mean square error (RMSE) for all the assets and the aggregate values by sectors according the Global industry Classification Standard (GICS).

Table 1.10: Giacomini and White test - *p-values*

	RMSE	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ
Financial									
HARRV	0.0239	-	-	-	-	-	-	-	-
HAR-RV-J	0.0216	0.06	-	-	-	-	-	-	-
HAR-FIGARCH	0.0229	0.00	0.90	-	-	-	-	-	-
HAR-CJ	0.0237	0.00	0.87	0.00	-	-	-	-	-
$RS^{+/-}$	0.0159	0.00	0.01	0.00	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0191	0.00	0.03	0.00	0.00	0.46	-	-	-
HAR-CJI	0.0119	0.00	0.00	0.00	0.00	0.00	0.00	-	-
LHAR-CJ	0.0127	0.00	0.03	0.00	0.00	0.00	0.01	0.17	-
LHAR-CJI	0.0063	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Industrial									
HARRV	0.0272	-	-	-	-	-	-	-	-
HAR-RV-J	0.017	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0247	0.00	1.00	-	-	-	-	-	-
HAR-CJ	0.0184	0.00	0.68	0.00	-	-	-	-	-
$RS^{+/-}$	0.0172	0.00	0.99	0.03	0.40	-	-	-	-
$RS^{J^{+/-}}$	0.0138	0.00	0.05	0.00	0.88	0.00	-	-	-
HAR-CJI	0.0153	0.00	0.54	0.00	0.05	0.01	0.85	-	-
LHAR-CJ	0.0141	0.00	0.04	0.00	0.10	0.01	0.31	0.01	-
LHAR-CJI	0.0083	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Energy									
HARRV	0.0146	-	-	-	-	-	-	-	-
HAR-RV-J	0.0167	0.03	-	-	-	-	-	-	-
HAR-FIGARCH	0.0207	0.00	0.86	-	-	-	-	-	-
HAR-CJ	0.0206	0.00	0.85	0.00	-	-	-	-	-
$RS^{+/-}$	0.0168	0.01	0.86	0.03	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0133	0.01	0.83	0.02	0.04	0.04	-	-	-
HAR-CJI	0.0133	0.47	0.11	0.00	0.05	0.01	0.80	-	-
LHAR-CJ	0.0125	0.17	0.31	0.00	0.05	0.01	0.09	0.10	-
LHAR-CJI	0.0096	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Materials									
HARRV	0.0204	-	-	-	-	-	-	-	-
HAR-RV-J	0.0202	0.02	-	-	-	-	-	-	-
HAR-FIGARCH	0.0222	0.08	0.97	-	-	-	-	-	-
HAR-CJ	0.0197	0.17	0.89	0.00	-	-	-	-	-
$RS^{+/-}$	0.0183	0.00	0.89	0.00	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0144	0.00	0.05	0.01	0.06	0.06	-	-	-
HAR-CJI	0.0159	0.00	0.74	0.00	0.07	0.03	0.84	-	-
LHAR-CJ	0.0125	0.00	0.01	0.00	0.01	0.06	0.07	0.06	-
LHAR-CJI	0.0114	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
Cons. Discr.									
HARRV	0.0563	-	-	-	-	-	-	-	-
HAR-RV-J	0.0262	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0516	0.00	0.97	-	-	-	-	-	-
HAR-CJ	0.0298	0.00	0.80	0.00	-	-	-	-	-
$RS^{+/-}$	0.0250	0.00	0.80	0.00	0.28	-	-	-	-
$RS^{J^{+/-}}$	0.0220	0.00	0.03	0.00	0.03	0.03	-	-	-
HAR-CJI	0.0229	0.00	0.03	0.00	0.00	0.00	0.44	-	-
LHAR-CJ	0.0195	0.02	0.00	0.00	0.00	0.00	0.40	0.08	-
LHAR-CJI	0.0129	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: The Table presents the average *p-values* across the assets belonging to the same sector of GW test, obtained using the out-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark.

Table 1.11: Giacomini and White test - *p-values* (cont'd)

	$RMSE$	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ
Health									
HARRV	0.0091	-	-	-	-	-	-	-	-
HAR-RV-J	0.0148	0.48	-	-	-	-	-	-	-
HAR-FIGARCH	0.0191	0.40	0.79	-	-	-	-	-	-
HAR-CJ	0.0156	0.30	0.41	0.20	-	-	-	-	-
$RS^{+/-}$	0.0138	0.33	0.06	0.02	0.09	-	-	-	-
$RS^{J^{+/-}}$	0.0100	0.98	0.00	0.00	0.03	0.04	-	-	-
HAR-CJI	0.0112	0.33	0.00	0.00	0.04	0.05	0.85	-	-
LHAR-CJ	0.0102	0.98	0.00	0.00	0.03	0.04	0.66	0.05	-
LHAR-CJI	0.0092	0.40	0.00	0.00	0.00	0.10	0.09	0.08	0.27
IT									
HARRV	0.0337	-	-	-	-	-	-	-	-
HAR-RV-J	0.0176	0.41	-	-	-	-	-	-	-
HAR-FIGARCH	0.032	0.00	0.75	-	-	-	-	-	-
HAR-CJ	0.0203	0.00	0.75	0.60	-	-	-	-	-
$RS^{+/-}$	0.0139	0.00	0.35	0.50	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0142	0.03	0.00	0.00	0.00	0.46	-	-	-
HAR-CJI	0.0135	0.00	0.00	0.31	0.29	0.23	0.25	-	-
LHAR-CJ	0.0138	0.00	0.00	0.00	0.34	0.27	0.30	0.23	-
LHAR-CJI	0.009	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Utilities									
HARRV	0.011	-	-	-	-	-	-	-	-
HAR-RV-J	0.0116	0.04	-	-	-	-	-	-	-
HAR-FIGARCH	0.0116	0.00	0.46	-	-	-	-	-	-
HAR-CJ	0.0121	0.00	0.37	0.40	-	-	-	-	-
$RS^{+/-}$	0.0104	0.00	0.31	0.31	0.76	-	-	-	-
$RS^{J^{+/-}}$	0.0081	0.00	0.03	0.00	0.00	0.00	-	-	-
HAR-CJI	0.0089	0.00	0.03	0.00	0.00	0.00	0.44	-	-
LHAR-CJ	0.0075	0.00	0.04	0.00	0.00	0.00	0.40	0.48	-
LHAR-CJI	0.007	0.00	0.04	0.00	0.00	0.00	0.40	0.04	0.02
Real Estate									
HARRV	0.0095	-	-	-	-	-	-	-	-
HAR-RV-J	0.0234	0.47	-	-	-	-	-	-	-
HAR-FIGARCH	0.0226	0.45	0.51	-	-	-	-	-	-
HAR-CJ	0.0222	0.45	0.46	0.49	-	-	-	-	-
$RS^{+/-}$	0.0228	0.46	0.42	0.41	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0095	0.19	0.33	0.00	0.03	0.01	-	-	-
HAR-CJI	0.0102	0.20	0.37	0.00	0.00	0.00	0.00	-	-
LHAR-CJ	0.0070	0.14	0.29	0.00	0.00	0.00	0.46	0.00	-
LHAR-CJI	0.0068	0.10	0.01	0.01	0.01	0.00	0.00	0.00	0.13
Cons. Stap.									
HARRV	0.0164	-	-	-	-	-	-	-	-
HAR-RV-J	0.0162	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0181	0.00	0.27	-	-	-	-	-	-
HAR-CJ	0.0176	0.02	0.24	0.33	-	-	-	-	-
$RS^{+/-}$	0.0153	0.02	0.08	0.00	0.09	-	-	-	-
$RS^{J^{+/-}}$	0.0093	0.01	0.00	0.00	0.08	0.08	-	-	-
HAR-CJI	0.0109	0.01	0.00	0.00	0.00	0.32	0.63	-	-
LHAR-CJ	0.0078	0.01	0.00	0.00	0.08	0.08	0.68	0.07	-
LHAR-CJI	0.0081	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54

Note: The Table presents the average *p-values* across the assets belonging to the same sector of GW test, obtained using the out-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark.

Table 1.12: MCS summary results

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ	LHAR-CJI
Financial									
JPM	-	-	-	-	-	-	-	-	1
BLK	-	-	-	-	-	-	-	2	1
BAC	-	-	-	-	-	-	-	-	1
AXP	-	-	-	-	-	-	-	-	1
WFC	-	-	-	-	3	-	-	2	1
Industrial									
ALK	-	-	-	-	2	-	-	-	1
EFX	-	-	-	-	-	-	-	2	1
FDX	-	-	-	-	-	-	-	-	1
UNP	-	-	-	-	-	-	-	-	1
KSU	-	-	-	-	3	-	-	2	1
Energy									
XOM	-	-	-	-	-	3	-	2	1
PXD	-	-	-	-	3	4	-	2	1
CVX	-	4	-	2	-	-	-	3	1
APA	3	-	-	-	2	1	-	-	-
EMN	3	-	-	-	-	-	-	1	2
Materials									
VMC	-	-	-	-	3	4	5	2	1
WRK	4	-	-	3	1	2	-	-	-
AVY	6	5	4	-	-	-	2	1	3
APD	-	-	-	-	4	3	-	1	2
EMN	5	6	-	-	4	3	5	2	1
Cons. Discr.									
RL	-	-	-	-	3	-	-	2	1
DIS	-	-	-	-	3	-	-	2	1
TSCO	-	-	-	-	-	-	-	-	1
RCL	-	-	-	-	-	-	-	-	1
CBS	-	-	-	-	2	-	-	-	1
Health									
BAX	1	8	-	5	7	2	4	6	3
ABT	3	-	-	6	2	3	5	4	1
JNJ	1	-	-	-	-	2	-	-	3
MDT	3	8	9	7	6	1	4	5	2
PFE	2	7	6	4	3	5	-	-	1
IT									
AAPL	-	-	-	3	4	-	2	4	1
EBAY	-	-	-	-	2	-	-	3	1
AMZN	-	-	-	-	-	-	-	-	1
INTC	-	-	-	-	3	-	4	2	1
ADBE	-	-	-	-	1	2	-	-	3
Utilities									
DUK	5	-	-	-	2	3	-	4	1
AEP	6	7	8	-	-	2	5	3	1
PPL	8	6	7	1	4	3	-	-	2
FE	5	-	-	1	4	2	-	-	3
DE	8	7	9	1	6	5	4	3	2
Real Estate									
BXP	6	5	-	-	4	2	-	3	1
FRT	6	7	-	4	5	2	-	3	1
VTR	-	-	-	3	2	-	-	1	-
PSA	6	-	-	5	3	4	-	2	1
SPG	-	-	-	4	5	3	-	2	1
Cons. Stap.									
PM	8	7	-	6	5	4	3	1	2
CVS	6	7	9	8	4	3	5	1	2
K	o	-	-	3	2	-	-	-	1
CL	-	5	6	4	3	2	-	-	1
KO	8	7	-	6	5	2	4	1	3

Note: The Table provides the summary of the Model Confidence Set (MCS) procedure recently developed by Hansen *et al.*(2011). Here, we show for each asset whether or not the model in column belongs to the set of the superior forecasting models. The sign (-) means the model does not belong to the set while the numbers indicate the rank produced by the loss function score. The confidence level is set at $\alpha = 0.2$.

Chapter 2

Hierarchical time varying estimation of pricing models

2.1 Summary

This paper presents a new hierarchical methodology for estimating multi factor dynamic asset pricing models. The approach is loosely based on the sequential Fama and MacBeth (1973) approach and developed in a kernel regression framework. However, our methodology uses very flexible bandwidth selection method which is able to emphasize recent data and information to derive the most appropriate estimates of risk premia and factor loadings at each point of time. The choice of bandwidths and weighting schemes, are achieved by a cross validation procedure; this leads to consistent estimators of the risk premia and factor loadings. Also, an out of sample forecasting exercise indicates that the hierarchical method leads to statistically significant improvement in forecast loss function measures, independently of the type of factor considered.

2.2 Introduction

The concept of a time varying risk premium is a standard idea in financial literature and many articles have approached the subject along the decades; e.g. Campbell and

Shiller (1988), Ferson and Harvey (1991) and Lewellen and Nagel (2006). Despite such evidence, the cornerstone method in empirical finance, by Fama and MacBeth (1973), estimates equity risk premia by a cross sectional regression method where the pricing of different types of risk factors are assumed constant.

This paper builds on Fama and MacBeth (1973) approach in the sense that a sequential hierarchical structure is developed in order to allow the risk factor estimates to change over time in a flexible yet tractable manner inside a kernel weighted regression framework. Our method maintains the Fama MacBeth (1973) stages of estimating risk factors, (or betas) from a time series regression and then the factor loadings (or gammas) from cross sectional regressions. However, we also include an additional stage for the selection of an optimal bandwidths via a cross validation procedure. Hence the main contribution of our methodology is to employ a flexible approach for the bandwidth selection, which essentially determines the speed of updates of the betas (risk factors) and also of the factor loadings, identifying an optimal time varying bandwidth level optimised for each assets at each point. This avoids imposing any priori structure and allows a natural data orientated way for incorporating economic and financial change, that is relevant for the pricing of assets under investigation. We refer to our approach as hierarchical since there is a clear hierarchy in the consideration of structural change by first allowing for time variation in the estimation of the parameters and then, in a second novel conceptual level, allowing for the bandwidth, determining the speed of change, to change itself. The method can also be seen as an extension of the least squares rolling window regression approach, which has extensively been used in empirical finance; e.g. Jagannathan and Wang (1996) and Lewellen and Nagel (2006).

Our empirical results overwhelmingly indicate the importance of removing the restriction of constant betas and the full superiority of our hierarchical methodology becomes apparent in terms of prediction of out of sample returns for a wide range of assets. Hence, our results show that time variation of risk associated with stocks and portfolios must be captured with an estimation procedure that on one hand avoids imposing excessive a priori structure and on the other takes into account the specific features of each asset and the time variation of its generating mechanism. Our methodology is indeed able to produce a increase in the forecasting performance greater between 4% to 7% than the alternative methods and independently for any type of model and asset.

The remainder of this paper is organized as follows: Section 2.3 provides a discussion of the contribution of this paper in light of the existing literature, introducing further the standard Fama and MacBeth (1973) approach. In Section 2.4 we delineate the hierarchical procedure introduced in the paper, discussing the estimation methodology. Further here, we formally present the link from our procedure, the static Fama and MacBeth (1973) and the other techniques proposed in literature. After introducing the data, Section 2.5, we finally illustrate in Section 2.6 the empirical application of our methodology, reporting also a series of robustness checks.

2.3 Background literature

The Capital Asset Pricing Model, or ($CAPM$) of Sharpe (1964), Lintner (1965) and Markowitz (1962) is a benchmark of asset pricing archetypes. The model implies that expected excess return on any asset is influenced by its sensitivity to the market, which is measured by the beta coefficient, times the market risk premia. Traditionally this beta is considered invariant over time and represents the covariance between the return of the asset and the return on the market portfolio. The basic model has been criticised by Black, Jensen and Scholes (1972), Fama and French (1992), Fama and MacBeth (1973) among others, on the grounds that only one factor, the market beta, is inadequate to describe the systemic risk. Hence many researchers have attempted to improve the basic $CAPM$ by the introduction of other factors. Most notably there is the three factor model of Fama and French (1993) which introduced the size, or SMB factor (positive returns are related to small size) and the high minus low, HML, factor (high book-to-market ratios are associated with higher returns). While Carhart (1997) introduced a fourth momentum factor, MOM, which describes the tendency of a stock price to continue recent trends. Other factors have been investigated in the literature, see Campbell (2000).

Further developments with extending the basic $CAPM$ have centred on implementing more flexible estimation strategies where the beta coefficient(s) are not necessarily assumed to be constant across time or space. For example, Harvey (1989), Ferson and Harvey (1991, 1993), Bollerslev, Engle and Wooldridge (1988) and Fama and French (1997, 2006), have suggested that a constant beta estimated using OLS does not capture the dynamics of the beta and is unable to satisfactorily explain the cross-section of average returns on equities.

Moving from such evidence, Adrian and Franzoni (2005) argue that models without time evolving betas fail to capture investor characteristics and may lead to inaccurate estimates of the true underlying risk. There are numerous factors that contribute to the variation in beta including regulation, economic and monetary policy, and exchange rates. Many researchers, such as Zolotoy (2011), show that variation in betas are more evident around important news announcements. Jagannathan and Wang (1996), Lettau and Ludvigson (2001), and Beach (2011) show that the conditional $CAPM$ with time varying beta generally outperforms an unconditional $CAPM$ with a constant beta.

One technique that is often used and embryonic version of the method used in this paper, is to take into account changes in the systematic risk of an asset through a rolling window OLS regression; e.g. Fama and MacBeth (1973) and Lewellen and Nagel (2006). While the former paper uses monthly returns over a five year window, the latter one employs returns at different horizons to capture the different rate of variation of risk over a variety of interval lengths (monthly, quarterly, semi-annually). Nevertheless, the main difficulty of rolling window regression approach is the attempt to capture local variation by having short intervals of data and that is incompatible with the desire of having tight standard errors and hence tight confidence intervals on the estimated beta parameters.

Other research has directly exploited the covariation between the market and other assets; see Engle (2002) and Bali and Engle (2010) who estimate time varying betas using multivariate dynamic conditional correlation methods to exploit correlations between cross sectional average returns of various factor portfolios. In line with realized measures, the usage of a realized beta allows for the instantaneous information adjustment. Further, realized betas allows for a flexible econometric framework that avoids fractional integration and/or co-integration between market variance and individual asset equity covariances with the market.

Again, another important way to handle the time variation in the beta coefficients is to model explicitly the evolution of the conditional distribution returns as a function of lagged state variables, see Jagannathan and Wang (1996), Ferson and Harvey (1999) and Adrian and Franzoni (2009) among the others. In all cases, the authors explicitly specify the covariance between the market and portfolio returns as affine functions of pre determined state variables. Jagannathan and Wang (1996), instead, develop a conditional version of the $CAPM$, augmented by a human-capital factor and show that it explains a substantial fraction of the cross-sectional variation in returns on 100 portfolios sorted on size and book-to-market. Further, Adrian and Franzoni (2009) allow for unobservable long-run changes in risk factor loadings, given by a learning process of rational in investors. The main drawback of these parametric approaches is that they require the correct specification for the functional form of the betas, or in other words they need to identify the right state variables. As pointed out by Ghysels (1998) and Harvey (2001), models with misspecified betas often feature larger pricing errors than models with constant betas.

Recently, new non parametric approaches have been proposed to overcome such limitations and arbitrary decisions. The idea is to assume that the parameter evolves smoothly over time and can be estimate it using a kernel weighted regression. Kernel methods to estimate time varying betas allows to use all the data efficiently in the estimation of conditional betas at any particular time. Ang and Kristensen (2012), introduce this methodology for investigating the asymptotic distributions for conditional and for long-run alphas and betas, averaged over time. For the choice of the bandwidth they suggest to use different bandwidths for conditional and long estimates in order for any finite-sample biases and variances to vanish. In addition, kernel smoothing estimators have the appealing feature that they nest, as a special case, rolling window estimates of betas (for example, Ferson and Harvey, 1991 and Petkova and Zhang, 2005; among many others).

Although, our analysis is mainly related with the evidence provided by Fama and MacBeth (1973) procedure (outlined in Section 2.1); it is important to cite the considerable contribution of Giraitis, Kapetanios and Yates (2014, 2018), Giraitis *et al.*(2015), and Giraitis *et al.* (2016). They provide a rigorous justification for using kernel methods to estimate structural change when the parameters, that undergo change, are not governed by a deterministic function of time, allowing for a wide class of stochastic processes that are characterised by persistence. This considerable relaxation of allowed processes is of importance since most economists and financial economists have strong prior that parameters change stochastically. We provide a further comparison of our methodology to the existing literature throughout the remainder of the paper.

2.3.1 Fama and MacBeth formulation

The seminal paper of Fama and MacBeth (1973), from the rest of the chapter $FMcB$, advocates a two step procedure to estimate risk premia in the multi factor asset pricing setting. The model assumes the coefficients are constant and estimates them using ordinary least squares regression. The first step regresses the excess risk free return of each asset, or portfolio, on various factors over time to determine the exposure of each factor and hence estimates the beta parameters. The second step consists of a cross section regression of the excess return of the assets against the factor exposures, or betas, at each point in time, in order to obtain a time series of risk premia coefficients, or gammas, for each factor. The great insight of Fama and MacBeth (1973) is to average these coefficients to obtain the expected premium for a unit of each risk factor and testing if these are adequately priced by the market.
In the seminal paper, they consider N assets and m factors; firstly the factor exposures, or betas are computed from the following time series regression produced for all the N assets:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{1,i}F_{1,t} + \beta_{2,i}F_{2,t} + ... + \beta_{m,i}F_{m,t} + u_{i,t},$$

with $i \in [1 : N]$, $t \in [1 : T]$, $R_{i,t}$ is defined as the nominal return on the i'th asset between period t and $t-1$ and $R_{f,t}$ denotes the risk free rate. Then $F_{j,t}$, with $j \in [1 : m]$, is a potential explanatory factor while $\beta_{j,i}$ represents the factor loading, that describes the degree of exposure of each asset to the factor, and $u_{i,t}$ is assumed to be $iid(0, \sigma_u^2)$.
The second step of the Fama and MacBeth (1973) method is to compute T cross section regressions of the excess return of the assets on the m estimated betas, $\hat{\beta}$, computed in the previous step. All these regressions use the same $\hat{\beta}$ since the objective of the Fama and MacBeth (1973) approach is to estimate the exposure of the N returns to the m factors loadings over time. Hence,

$$(R_{i,t} - R_{f,t}) = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{1,i} + \gamma_{2,t}\hat{\beta}_{2,i} + ... + \gamma_{m,t}\hat{\beta}_{m,i} + \varepsilon_{i,t},$$

where γ_js measure the risk premia associated with each F_j. Hence the method determines $m+1$ series of the γs, which are also generally considered to be constant. If the model is well specified and all the factors considered are significant, then the risk loadings explain cross sectional differences, $\hat{\bar{\gamma}}_0 = 0$ and $\hat{\bar{\gamma}}_j$ represent the average risk

premia for each factor.
The main limitation of this the two-pass cross-sectional method is being subject to the error-in-variables (EIV) problem, due the use of estimated betas in the second step. Indeed, although the procedure produce consistent risk premia estimators, as the time-series sample size tends to infinity and the cross section size is fixed, the traditional $FMcB$ standard errors are not consistent, requiring an asymptotic bias correction. In other words, with higher errors in the estimation of beta the EIV issues becomes more severe in the sense that the ex-post risk premia estimator itself is inconsistent.
Recently, Adrian, Crump and Moench (2015) use the weighted kernel estimator by Ang and Kristensen (2012), to propose a methodology more robust to misspecification errors. Their empirical application features good pricing properties across stocks and bonds and shows notable time variation of expected returns associated with highly significant dynamic price of risk parameters. Moreover, they show that Gaussian kernel estimator yields smaller pricing errors than simple rolling window regressions for both specifications with constant and time-varying prices of risk.

2.4 Hierarchical methodology

The main contribution is to develop a flexible methodology, inside the kernel regression framework, to easily allow time variation in both the betas and gammas of the baseline Fama and MacBeth (1973) approach.
The main tool to achieve this is to have a flexible bandwidth parameter which essentially controls the weight given to local information for updating the beta and gamma coefficients. The innovations, here, are: optimising the bandwidth selection based on out of sample cross validation methods and allowing the bandwidth to change over time. As outlined before, there has already been considerable awareness of the importance of incorporating time variation in the estimation of the multi factor models but less in the computation of the bandwidth. For example, Ang and Kristensen (2012) and Adrian, Crump and Moench (2015), both use an *a priori* constant bandwidth, selected according data characteristics. In order to add a degree of freedom to this process, we use a cross validation procedure for identifying an optimal bandwidth level optimised for each assets at each point in time, for deriving improved local estimates

of the betas. In other words we extends not only Fama and MacBeth (1973) but also Adrian, Crump and Moench (2015) to the case where all the parameters are time varying, including the bandwidth.
Hence, the novelty of the approach is to identify an optimal time varying bandwidth for each asset that helps to reduce the forecast errors of the risk premia via a more accurate estimation of the factor loadings. The procedure can be described as follow. In the first step, we use a cross validation approach to identify the optimal bandwidth asset specific. In second step, asset returns are regressed in the time series on risk factors, using the bandwidths obtained before, generating the time varying risk betas for each assets. In the final step, the price of risk parameters are computed by regressing the excess return on the betas from the time series regression, cross sectionally.

2.4.1 Cross validation - bandwidth choice

As previously mentioned, an important aspect of our paper is the use of cross validation to search for the most appropriate bandwidth in the kernel function that sets the degree of smoothness of the estimates. This parameter turns out to be critical in providing the appropriate degree of persistence in determining the *memory* of the window used for the estimation of the time varying coefficient of the model. Following previous literature, we decide to consider here the classical 3 factors model proposed by Fama and French (1995) model: market factor, $MRKT$, size factor, SML, and book-to-market factor, HML.
The first part of our hierarchical approach is to calculate the time varying parameters (TVP) associated with the coefficients of the factors (βs). The method, we use, is based on a kernel weighted regression, hence

$$(R_{i,t} - R_{f,t})_h = \beta_{1,t,i,h} F_{MRKT,t} + \beta_{2,i,t,h} F_{SMB,t} + \beta_{3,i,t,h} F_{HML,t} + u_{i,t,h}, \tag{2.1}$$

with $i \in [1:N]$ number of assets, $t \in [1:T]$ period of time, $k \in [1:3]$ number factors and h is the bandwidth parameter, to be discussed later, such that $h \in [0.05, 0.9]$ with an interval of 0.05. Further, it is generally assumed throughout the paper that $u_{n,t+1}$ is $i.i.d.(0, \sigma^2)$. The β parameters are estimated by an extension of the methodology of Giraitis, Kapetanios and Yates (2014), summarized in the Appendix E of this paper.

Hence the beta for the k^{th} factor is estimated by

$$\hat{\beta}_{k,t,i,h} = \frac{\sum_{t=1}^{T} K(\frac{t-j}{T})\,(R_{i,t} - R_{f,t})\,F_{k,t}}{\sum_{t=1}^{T} K(\frac{t-j}{H})F_{k,t}^{2}}, \tag{2.2}$$

where $K(\frac{t-j}{H})$ is assumed to be a Gaussian kernel function. The bandwidth, H, represents the degree of smoothness of the estimates. In other terms, if the bandwidth is small, the estimates will be under smoothed, with high variability, otherwise if the value of H is big, the resulting estimators will be over smooth and farther from the real function. Different approaches have been proposed to handle the choice of the bandwidth. Ang and Kristensen (2012) suggest to optimise the choice of the bandwidth for conditional and long estimates in order to reduce any finite-sample biases and variances. Giraitis, Kapetanios and Yates (2014, 2018), instead, proved under very mild condition that if the bandwidth is $H = T^h$, with the bandwidth parameter $h = 0.5$, the estimator shows desirable properties such as consistency and asymptotic normality and in addition provide valid standard errors.

Our method, instead, is agnostic on the choice of the parameter h and then on the bandwidth. We decide to include an additional liberalization of the parameter, neglecting a common bandwidth assumption and allowing it to vary across time and assets. Specifically, we use a cross validation procedure for identifying a time varying bandwidth optimised for each asset. In other words, we produce an optimal parameter $h_{i,t}^{opt}$, for each asset and time period, selected from an out of sample, one-step ahead forecasting comparison over a grid search of h which incorporates 18 different values of h, for the grid of $h \in [0.05; 0.9]$ with an interval of 0.05 for each grid.

For the remaining of the paper when we talk about optimisation of the bandwidth, we mean the choice of the parameter h inside the bandwidth formula, $H = T^h$.

The end of this stage, the process generates for each asset, i, a time series of beta estimates for different values of the bandwidth parameter h. These estimated betas allow the identification of the price of risk factor loadings for different values of h, γ_h; using the following equation:

$$(R_{i,t} - R_{f,t})_h = \gamma_{0,h} + \widehat{\beta}'_{1,t,i,h}\gamma_{1,t,h} + \widehat{\beta}'_{2,t,i,h}\gamma_{2,t,h} + \widehat{\beta}'_{3,t,i,h}\gamma_{3,t,h} + \varepsilon_{n,t,h}, \tag{2.3}$$

where $\varepsilon_{n,t,h}$ is assumed to be $i.i.d.(0, \sigma_\varepsilon^2)$. This process generates $k+1$ series of γs (including the constant) for every value of the bandwidth parameter, h. The cross validation procedure then compares the forecasting performance of the competing models, via the computation of the forecast errors $e_{i,t+1,h}$. The initial T_0 observations are the *training period*, while the remaining ones, $(T - T_0)$, define the out of sample period. The training period is fixed at 60 observations, or 5 years of data, we also performed robustness tests with different values of T_0 of 120 and 180. The one step ahead forecast for each asset is then obtained from the following regression

$$(\widehat{R_{i,t+1} - R_{f,t+1}})_h = \widehat{\gamma}_{0,h} + \widehat{\beta}'_{1,t,i,h}\widehat{\gamma}_{1,t,h} + \widehat{\beta}'_{2,t,i,h}\widehat{\gamma}_{2,t,h} + \widehat{\beta}'_{3,t,i,h}\widehat{\gamma}_{3,t,h}. \tag{2.4}$$

The forecast errors, $e_{i,t+1,h}$ are computed for each period and for each of the eighteen different values of h. The time varying $RMSE$ is calculated at each point in time and for each asset; and the value of h is chosen via a minimization procedure. Several different criteria and approaches were investigated to compute this measure; including rolling window, and non parametric kernel smoothed technique. The former approach refers to the classical rolling window method with different window, w such that $w \in [12; 24]$. Hence the unadjusted rolling $RMSE$ is given by

$$RMSE_t^{roll} = \sqrt{\frac{1}{w}\sum_{j=1}^{w} e_{i,t+j,h}^2}, \tag{2.5}$$

while the kernel weighted $RMSE$ is computed instead:

$$RMSE_t^{kern} = \sqrt{\sum_{j=1}^{T} W\left(\frac{t-j}{H^{(i)}}\right) e_{i,t+j,h}^2}, \tag{2.6}$$

with $H^{(i)} = T^{h'}$ and $h' \in [0.05; 0.9]$. Clearly when $W(H) = 1$ the formula reduces to the regular $RMSE_t$ formula in Equation (2.5). Both approaches generates a matrix of 18 columns and $(T - T_0 - w)$ or $(T - T_0)$ rows according to the method used, for each asset. This matrix of $RMSE_t$ is then used to determine the optimal values of h for each asset, $h_{i,t}^{opt}$, such as the value that produces the lowest $RMSE$. The approach generates a time series of optimal values of h, that are used in the second step of our procedure to get a more accurate estimation of βs coefficient.

2.4.2 Estimation of factor risk loadings

Once obtained the matrix with the optimal values of h, we use it to compute time varying factor risk loading. The point of our procedure is to allow a full liberalization of the parameters, optimizing the choice of a time varying bandwidth for each asset. Hence, we compare the forecasting performance of our method with different approaches for the computation of β:

(i) Classical Fama and MacBeth (1973) approach where the betas are computed using a, OLS rolling window approach with five years window.

(ii) A kernel weighted approach with $h = 0.5$. As showed by Giraitis *et al.* (2014), this bandwidth allows to get smooth estimates with desirable properties such as consistency and asymptotic normality and in addition provides asymptotically valid standard errors. This model is then used as a benchmark.

(iii) Alternative kernel approach, where h is fixed for each assets and time and is determined from a poll average of the optimal bandwidth parameters, $h_{i,t}^{opt}$, as follow:

$$\bar{h}^{Polling} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} h_{t,i}. \tag{2.7}$$

(iv) A further kernel regression approach, with h computed by averaging the optimal bandwidth parameters across assets. While the parameter varies over time, it is not asset specific:

$$\bar{h}_t^{Average} = (N)^{-1} \sum_{i=1}^{N} h_{t,i}. \tag{2.8}$$

(v) A kernel approach that uses the optimal $h_{i,t}^{opt}$, which are different for each asset to give; this method will be named *Specific*.

All these 5 approaches are implemented in the three factor Fama and French (1993) model:

$$R_{i,t} - R_{f,t} = \beta_{1,t,i,m} F_{MRKT,t} + \beta_{2,i,t,m} F_{SMB,t} + \beta_{3,i,t,m} F_{HML,t} + u_{i,t,m}, \tag{2.9}$$

where $m \in [1:5]$ represents one of the aforementioned approaches used for the computation of the factor loadings, βs. The coefficients are computed according Equation (2.2).

2.4.3 Estimation of risk premia

The time varying estimates of beta, $\hat{\beta}_t$, are then used in the third step of our procedure, which facilitates computation of risk premia associated with the factors under investigations, γs. Hence this stage of our procedure has some similarities with Fama and MacBeth (1973) article, except that at each point in time, we consider multiple time varying estimates of beta instead of fixed as a constant. Our hierarchical methodology then replaces the assets' excess returns by their corresponding time varying kernel weighted average, $(R_{n,t} - R_{f,t})$, for coherence in terms of degree of smoothness. Indeed, they are computed using the bandwidth, h that has been selected in the previous step for the computation of the βs. The kernel weighted averages for the excess returns are then

$$(\widehat{R_{i,} - R_{f,t+1}}) = \sum_{k=1}^{T} K\left(\frac{t-k}{H^*}\right)(R_{i,k} - R_{f,k}), \tag{2.10}$$

where $K(\frac{t-k}{H^*})$ is the same continuously bounded kernel function and $H^* = T^{h^m}$ with h^m that identifies the bandwidth used at the previous step for the computation of the coefficients, $m \in [1:5]$. These smoothed excess returns are then used for the OLS regressions, to identify the risk premia,

$$(\widehat{R_{i,t} - R}_{f,t}) = \gamma_{0,m} + \widehat{\beta}_{1,t,i,m}\gamma_{1,t,m} + \widehat{\beta}_{2,t,i,m}\gamma_{2,t,m} + \widehat{\beta}_{3,t,i,m}\gamma_{3,t,m} + \varepsilon_{i,t,m}. \tag{2.11}$$

This results in $m+1$ series of $\hat{\gamma}$ (including the constant), for each of the five different approaches previously considered for the estimation of the βs.
The last stage of the hierarchical approach is to select the best methodology in terms of $RMSE$ minimization for an out of sample forecasting exercise. To do that, we firstly forecast the average excess return across all assets using the average of the estimated betas, which realizes the time series of forecasts of the average,

$$\overline{R_{t+1} - R_{f,t+1}} = \hat{\gamma}_{0,m} + \hat{\gamma}_{1,t,m}\bar{\hat{\beta}}_{1,t,m} + \hat{\gamma}_{2,t,m}\bar{\hat{\beta}}_{2,t,m} + \hat{\gamma}_{3,t,m}\bar{\hat{\beta}}_{3,t,m}, \tag{2.12}$$

where

$$\overline{R_{t+1} - R_{f,t+1}} = \frac{1}{N}\sum_{i=1}^{N}(R_{t+1} - R_{f,t+1}), \tag{2.13}$$

and

$$\bar{\beta}_{j,t,m} = \frac{1}{N}\sum_{i=1}^{N}\beta_{j,t,i,m}.$$

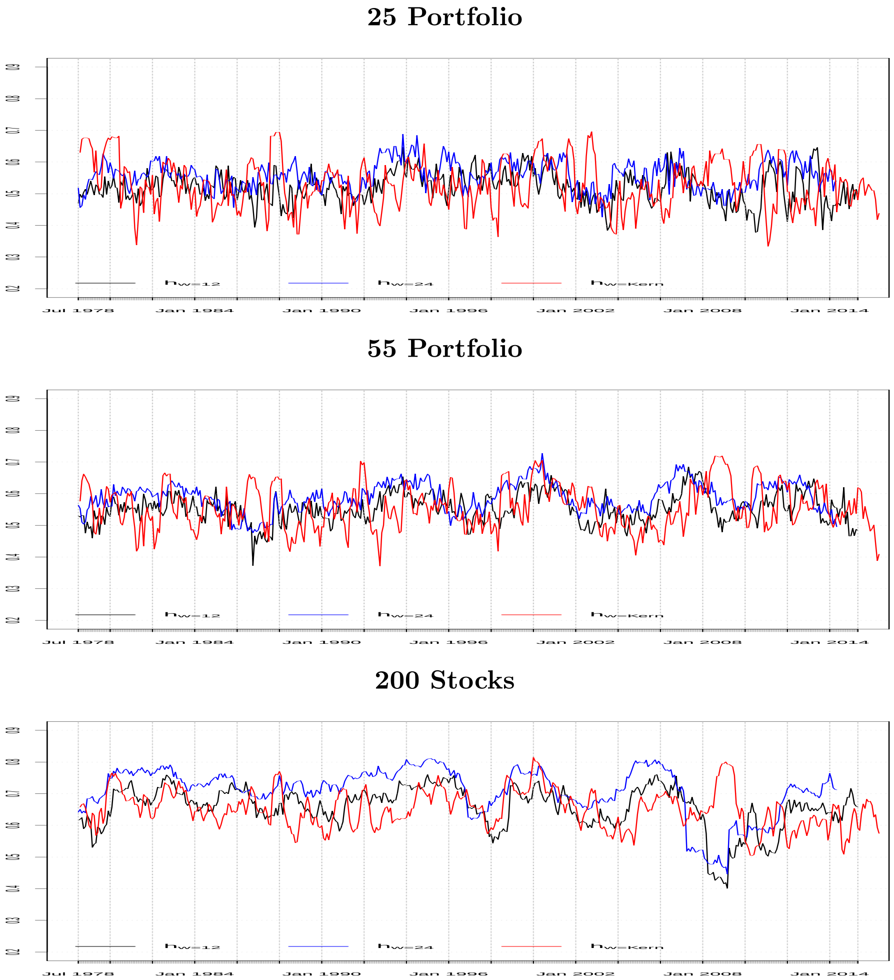
Then, the $RMSE$ are then computed for each method and compared to identify the estimation method, using the Diebold and Mariano (1995) test.

2.5 Data

The new hierarchical methodology is now applied to three different financial returns data sets. The first contains $N = 25$ sized portfolios sorted by size and book-to-market, while the second one $N = 55$; 25 from the aforementioned dataset and 30 sized portfolios sorted by industry, both available from Ken French's on-line data library. We further use 200 Standard & Poor's constituents from the Center for Research in Securities Prices (CRSP), so that $N = 200$. We consider excess return over the 30-day Treasury bill yield, with the total series covering the period from August 1973 through May 2016, for a total of $T = 514$ observations available again from Kenneth French's on-line web site.

We use the following set of factors in our subsequent analysis; namely the excess return on the market, $MRKT$, value-weight return of all $CRSP$ firms incorporated in the US and listed on either the $NYSE$, $AMEX$, or the $NASDAQ$. The small minus big, SMB, factor and the high minus low, HML factors are derived in the same way as in Fama and French (1993) and are available from Ken French's on-line data library[1].

[1]In Section 2.6.1 and Appendix F, we further consider, as robustness check, other well know financial factors: the momentum factor, MOM, by Carhart (1997), which is computed as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. Finally our methodology has also been tested on the five factor model from Fama and French (2015). These additional factors represent the *robust minus weak*, RMW and the *conservative minus aggressive*, CMA factors. The RMW factor is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolio, while CMA represents the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios. All the aforementioned factors are available from Ken French's on-line data library.



Note: The Figure reports the plots of the optimal bandwidth parameters considering different datasets and methods for computing the $RMSE_t$, as discussed in Section 2.4.1. The bandwidths reported in red are computed according Equation (2.5) using the classical rolling window approach with $w = 12$, while in blue with $w = 24$. The bandwidths in black instead are computed using the kernel average method discussed in Equation (2.6).

FIGURE 2.1: Time Varying optimal bandwidth parameters

2.6 Empirical results of the hierarchical analysis

Following the details of the above methodological framework, Table 2.1 provides the descriptive statistics for the optimal bandwidth parameters, h^{opt}, for all the different data sets. The results are categorized in terms of the methodology used to compute the time varying $RMSE_t$ measure: $h^{opt}_{w=12}$ and $h^{opt}_{w=24}$ refer to the conventional rolling window approach with windows of 12 and 24 observations, Equation (2.5), while h^{opt}_{kern}

Table 2.1: Descriptive Statistics for the optimal bandwidth parameter

	Obs.	**Mean**	**St. Dev.**	**Min**	**Max**	**Skew**	**Kurt**
25 Portfolio							
$h_{w=12}$	443	0.51975	0.04937	0.37800	0.64600	-0.03670	-0.04397
$h_{w=24}$	431	0.55826	0.04407	0.42600	0.68800	-0.01419	-0.32630
h_{Kern}	454	0.52804	0.07684	0.33400	0.69600	0.11501	-0.55003
55 Portfolio							
$h_{w=12}$	443	0.55331	0.04655	0.37273	0.68455	0.05771	0.02067
$h_{w=24}$	431	0.58815	0.04586	0.47636	0.72727	-0.05941	-0.35675
h_{Kern}	454	0.55532	0.06884	0.37182	0.71909	0.23544	-0.48401
200 Stocks							
$h_{w=12}$	443	0.65671	0.06742	0.40100	0.76000	-1.22993	1.83550
$h_{w=24}$	431	0.70924	0.05967	0.44625	0.81100	-1.31970	1.52070
h_{Kern}	454	0.65222	0.07571	0.50425	0.81475	0.14502	-0.16497

Note: The Table reports the descriptive statistics of the optimal bandwidth parameters considering different datasets (25 portfolios, 55 portfolios or individual stocks, 200) and obtained using different methods for computing the $RMSE_t$ as discussed in Section 2.4.1: the former method refers to the classical rolling window approach with different w such that $w \in [12; 24]$, Equation (2.5), while the latter involves a kernel average method, Equation (2.6).

refers to the kernel approach, Equation (2.6). From the analysis of the panels containing the portfolio results, it can be seen that the cross validation procedure is remarkably consistent in choosing h near 0.50 and standard deviation of the estimates relatively small lying in the range 0.044 to 0.076 for all the methodologies. Regular t-tests were unable to reject the hypothesis that $h = 0.50$ for any of the portfolio classifications. This finding is particularly interesting since $h = 0.50$ is the theoretical value identified by Giraitis, Kapetanios and Yates (2014) as being the optimal value for h in terms of achieving an appropriate rate of convergence to an asymptotic distribution of the TVP. However, the averages for h^{opt} for the individual stocks data are higher than the ones for the two portfolios, being around $h = 0.65$. This can be interpreted as the need to increase the degree of smoothness when we use data with an high level of heterogeneity as the dataset containing the constituents of $S\&P$ 500. Further, the analysis of across the different methods for computing the $RMSE_t$ shows that the non parametric *kernel* approach provides the highest values for the standard deviations for each portfolio.

Figure 2.1 plots the selected optimal bandwidth parameters, averaged across assets as in Equation (2.8), for each of the different methodology for computing the time varying $RMSE_t$ and also for different data sets. All the methods provide an erratic mean reverting path, centred around 0.5, where the kernel approach confirms to be the most volatile in all the data combination. In general, the non parametric approach appears to be the most volatile; and is the only one that increases in the Global Financial Crisis, GFC.

TABLE 2.2: Factor risk loading estimates for Ford - Stocks

	Constant βs			*Rolling*			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.7331	0.6690	0.8870	0.6008	0.7378	0.5933	0.6404	0.7413	0.1441
	(0.1164)	(0.1728)	(0.1744)	(0.2993)	(0.4719)	(0.4822)	(0.1148)	(0.2719)	(0.2926)
$h_{w=24}$	0.7379	0.6988	0.8895	0.6249	0.7443	0.6284	0.6435	0.7615	0.1341
	(0.1164)	(0.1728)	(0.1744)	(0.2778)	(0.4361)	(0.4428)	(0.1199)	(0.2864)	(0.3057)
h_{Kern}	0.7449	0.6571	0.8763	0.5719	0.7405	0.5524	0.6460	0.7135	0.6401
	(0.1164)	(0.1728)	(0.1744)	(0.3260)	(0.5146)	(0.5292)	(0.1129)	(0.2622)	(0.2848)
	h Pooling			*Average h*			*h Specific*		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.7267	0.7116	0.7683	0.7375	0.7060	0.7757	0.6333	0.7759	0.8200
	(0.0438)	(0.1041)	(0.1067)	(0.0466)	(0.1217)	(0.1176)	(0.0482)	(0.0896)	(0.0916)
$h_{w=24}$	0.7410	0.7150	0.7986	0.7804	0.6880	0.8053	0.6197	0.7704	0.8294
	(0.0343)	(0.0742)	(0.0782)	(0.0376)	(0.0964)	(0.0921)	(0.0401)	(0.0679)	(0.0828)
h_{Kern}	0.7214	0.6726	0.7392	0.7105	0.6767	0.7621	0.6441	0.6066	0.7476
	(0.0438)	(0.1020)	(0.1067)	(0.0471)	(0.1163)	(0.1212)	(0.0721)	(0.1281)	(0.1505)

Note: Average estimates of factor risk loadings for Ford stock using the 200 stocks dataset for the computation of the optimal bandwidth. There are 6 different methodology: simple ordinary least square regression (*Constant*), *Rolling* window approach (with a 5 years window), and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time, Equation (2.7); *Average*, a unique time varying bandwidth coming from the average of h across asset, Equation (2.8); *Specific*, multiple time varying bandwidths, one for each asset and time. In parenthesis, there are the averages of the standard errors.

Tables 2.2 to 2.3 provide details of the estimated beta coefficients for representative assets for each dataset. For the portfolio datasets we consider the estimates for the median portfolios, named $ME3.BM3$, while for the for the constituents of the $S\&P500$, we analyse the $Ford$ index. In each panel, we also provide the standard errors for the each of the three factor loadings: $MRKT$, SMB and HML. The estimated market beta, $\hat{\beta}_{MRKT}$, is close to the unity for all the portfolio datasets while it is around 0.7 for the Ford stock, in line with the literature. The standard errors provided by the $Specific$ approach are the smallest and are very important for subsequent efficient estimation of risk premia[2]. Figure 2.2 presents the factor risk loading estimates. Each one of the nine separate panels shows five TVP beta estimates derived from the methodologies presented in Section 2.4. In particular, the rolling window is displayed with a green line, the kernel estimate with constant bandwidth parameter of $h = 0.5$ in black, the constant bandwidth parameter from poll average, $Polling$, in purple, further, the time varying h set equal for all the asset, $Average$, with a blue line and finally the time varying h optimised for each asset, $Specific$, with a red line. The last three methods all use the Gaussian kernel.

In all the scenarios the time varying estimates are centred around the constant ones highlighting the correctness of the methodology. Further, we note a similar path for all the kernel estimates with the ones produced by the classical rolling window approach. In accordance with Adrian, Crump and Moench (2015), we observe that the estimates produced with the classical approach exhibit overall an higher variation than the ones produced with kernel approaches. Very peculiar is the path shown by the estimates computed according the $Specific$ approach. Although, these estimates follow a path in line with the others, they are characterised by numerous sudden changes along the sample period. These changes appear to be asset specific and hence different asset by asset. A rigorous analysis of all the changes has been produced, and we were able to identify a valid explanation according economical, political and financial events for most of the stock assets. For portfolio data, instead, such procedure was harder to perform given the nature of the dataset, but such argument is beyond the aim of the paper[3].

[2]From now on, throughout the remainder of the paper, we exclusively report the results concerning the kernel approach methodology for the computation of the time varying $RMSE_t$, as in Equation (2.6). The results for the other two approaches are available upon request to the authors.

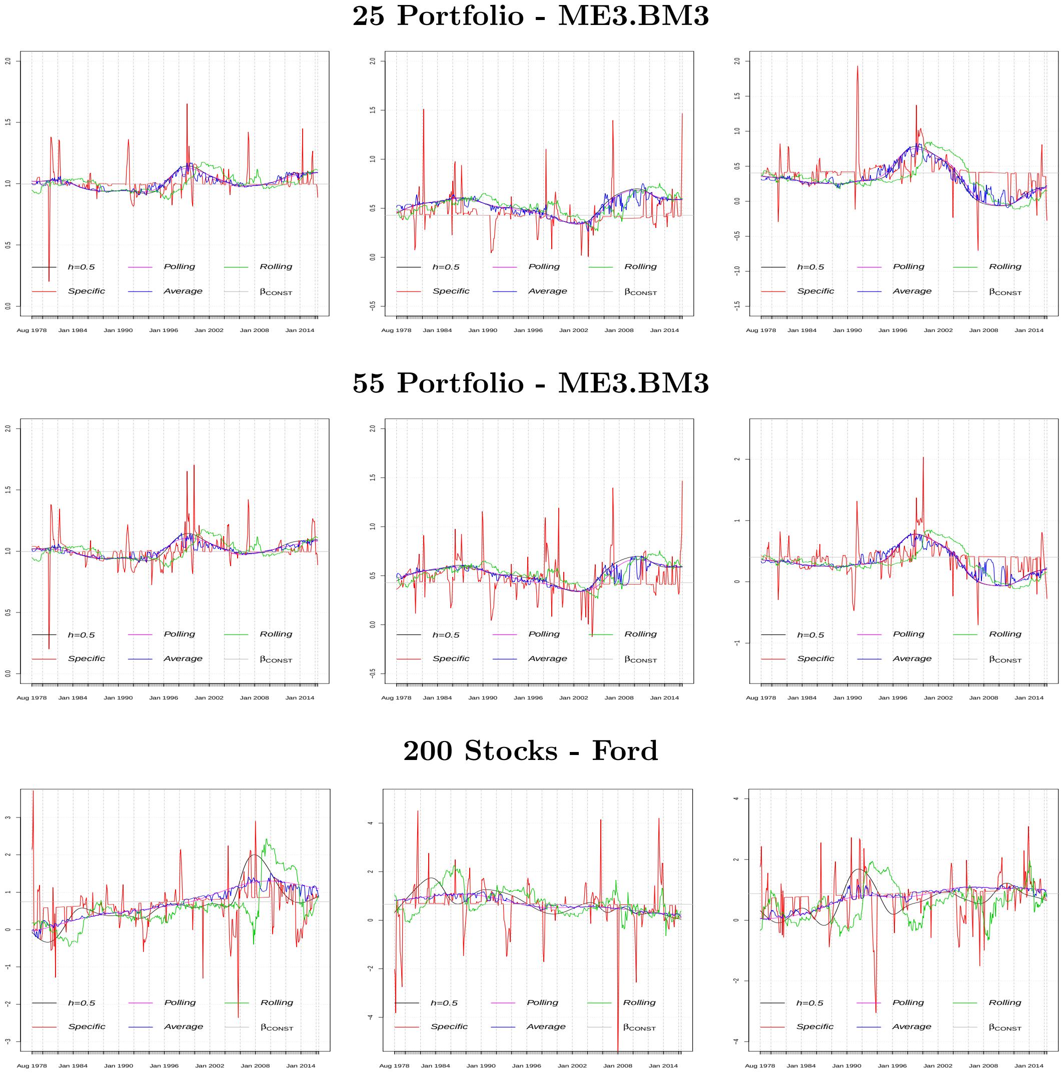
[3]Further details are available upon request to the authors.

TABLE 2.3: Factor risk loading estimates for ME3.BM3 - 25Portfolio

Panel A	*Constant βs*			*Rolling*			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.9943	0.4265	0.4017	1.0034	0.5275	0.3101	1.0142	0.5363	0.3035
	(0.0175)	(0.0259)	(0.0262)	(0.0413)	(0.0619)	(0.0651)	(0.0119)	(0.0282)	(0.0303)
$h_{w=24}$	0.9935	0.4211	0.4059	1.0022	0.5230	0.3165	1.0138	0.5342	0.2962
	(0.0175)	(0.0259)	(0.0262)	(0.0384)	(0.0571)	(0.0596)	(0.0122)	(0.0291)	(0.0311)
h_{Kern}	0.9969	0.4282	0.4029	1.0052	0.5296	0.3036	1.0135	0.5304	0.3024
	(0.0175)	(0.0259)	(0.0262)	(0.0453)	(0.0681)	(0.0721)	(0.0120)	(0.0279)	(0.0303)
	h pooling			*Average h*			*h specific*		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	1.0132	0.5322	0.3072	1.0114	0.5281	0.3085	1.0123	0.4670	0.3830
	(0.0104)	(0.0249)	(0.0265)	(0.0109)	(0.0267)	(0.0289)	(0.0048)	(0.0103)	(0.0158)
$h_{w=24}$	1.0099	0.5204	0.3101	1.0080	0.5200	0.3101	1.0041	0.4684	0.3792
	(0.0083)	(0.0203)	(0.0208)	(0.0086)	(0.0217)	(0.0221)	(0.0053)	(0.0080)	(0.0167)
h_{Kern}	1.0116	0.5241	0.3074	1.0089	0.5199	0.3067	1.0126	0.4461	0.4163
	(0.0099)	(0.0233)	(0.0250)	(0.0115)	(0.0268)	(0.0304)	(0.0067)	(0.0165)	(0.0176)

Panel B	*Constant βs*			*Rolling*			$h = 0.5$		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	0.9943	0.4265	0.4017	1.0034	0.5275	0.3101	1.0142	0.5363	0.3035
	(0.0175)	(0.0259)	(0.0262)	(0.0413)	(0.0619)	(0.0651)	(0.0119)	(0.0282)	(0.0303)
$h_{w=24}$	0.9935	0.4211	0.4059	1.0022	0.5230	0.3165	1.0138	0.5342	0.2962
	(0.0175)	(0.0259)	(0.0262)	(0.0384)	(0.0571)	(0.0596)	(0.0122)	(0.0291)	(0.0311)
h_{Kern}	0.9969	0.4282	0.4029	1.0052	0.5296	0.3036	1.0135	0.5304	0.3024
	(0.0175)	(0.0259)	(0.0262)	(0.0453)	(0.0681)	(0.0721)	(0.0120)	(0.0279)	(0.0303)
	h pooling			*Average h*			*h specific*		
	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}	β_{MRKT}	β_{SMB}	β_{HML}
$h_{w=12}$	1.0105	0.5234	0.3145	1.0093	0.5192	0.3221	1.0078	0.4694	0.3856
	(0.0084)	(0.0203)	(0.0212)	(0.0088)	(0.0214)	(0.0231)	(0.0078)	(0.0132)	(0.0204)
$h_{w=24}$	1.0068	0.5104	0.3203	1.0051	0.5069	0.3277	1.0071	0.4741	0.3925
	(0.0069)	(0.0169)	(0.0171)	(0.0072)	(0.0179)	(0.0190)	(0.0071)	(0.0125)	(0.0219)
h_{Kern}	1.0091	0.5164	0.3133	1.0071	0.5105	0.3203	1.0058	0.4703	0.3864
	(0.0083)	(0.0197)	(0.0209)	(0.0095)	(0.0217)	(0.0242)	(0.0107)	(0.0200)	(0.0237)

Note: Average estimates of factor risk loadings for the portfolio ME3.BM3 using 25 (Panel) and 55 Portfolio for the computation of the optimal bandwidth, respectively in Panel A and Panel B. There are 6 different methodologies: simple ordinary least square regression (*Constant*), *Rolling* window approach (with a 5 years window), and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time, Equation (2.7); *Average*, a unique time varying bandwidth coming from the average of h across asset, Equation (2.8); *Specific*, multiple time varying bandwidths, one for each asset and time. In parenthesis, there are the averages of the standard errors.



(a) β_{SMB} (b) β_{SMB} (c) β_{HML}

Note: The figure provides the estimates of factor risk loadings computed using the normal approach *Constant* β (grey line), *Rolling* window, with 5 years estimation period (green line) and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5* (black line); *Polling* a single value of h coming from the poll average the cross asset and time as shown in Equation (2.7), (purple line); *Average*, a unique time varying bandwidth coming from the average of h across asset, Equation (2.8), (blue line); *Specific*, multiple time varying bandwidths, one for each asset and time (red line). The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach, as discussed in Equation (2.6).

FIGURE 2.2: A dynamic comparison of the factor loading estimates

Moreover, the different degrees of variation of the methods give rise of interesting observation. As expected, the beta estimates for portfolio datasets exhibit a lower degree of variation than those that employ stock indexes. This is presumably due to noise using stock data and the loss of information induced by grouping stocks to build a portfolio (Lo and MacKinlay, 1990). In general, the betas on the MKT and HML factors are the one that most often switch sign, while the SMB appear to be the most stable factor. Table 2.4 provides estimates and the respective standard errors of the risk premia parameters, γ_i with $i \in [0, 3]$, including also the constant term. The Newey West standard errors are also displayed in the last column. Further, it presents results for h^{opt}, computed using $RMSE_t$ with kernel averaging approach. The results for the other 2 parametric approaches are available online.

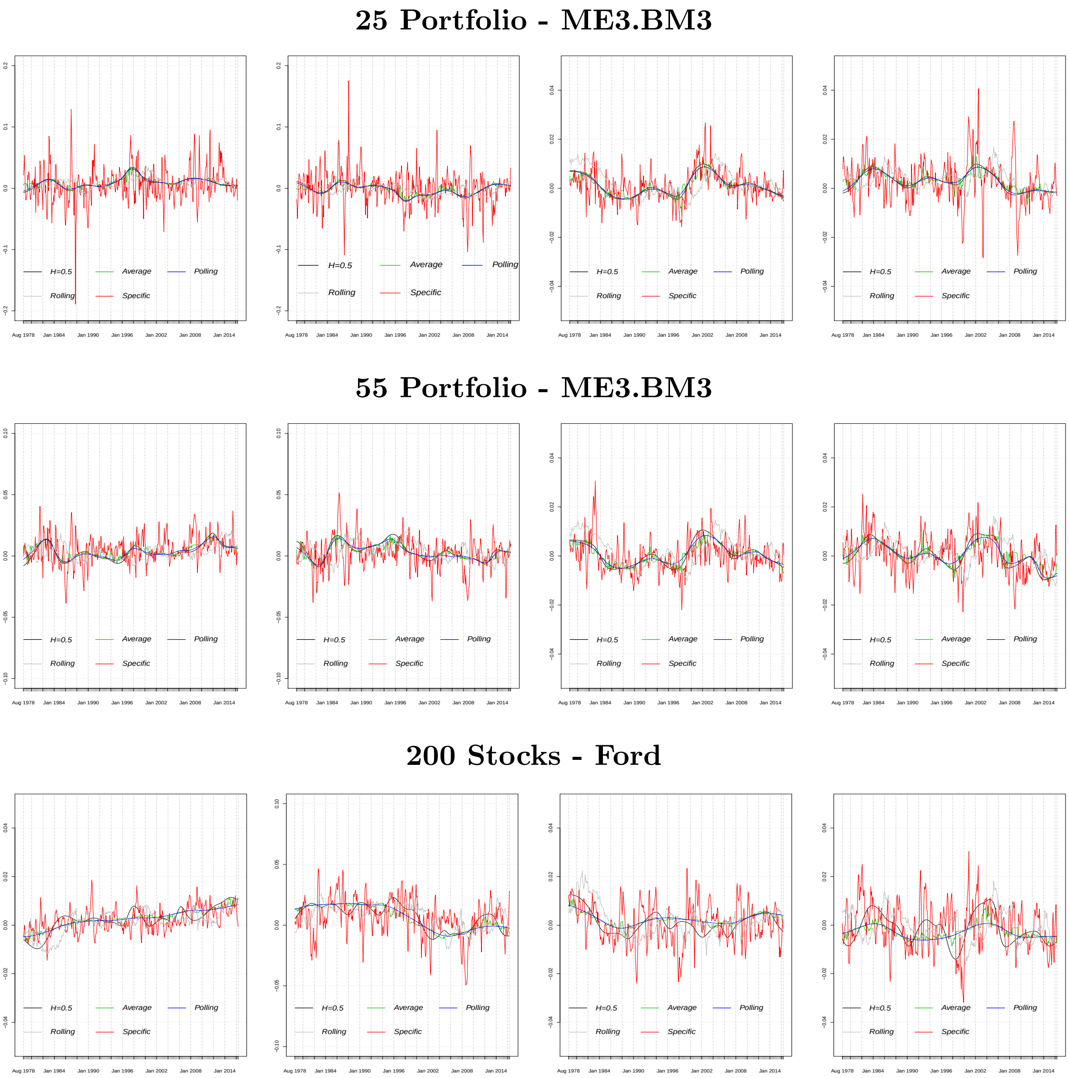
The average prices of risk appear to be very similar across the different methods and within each dataset. The $Specific$ method shows the smallest standard errors despite the sample considered. The sample size appears to matter and affects the significance of the price of all the factors. In particular, SMB is priced only considering individual stocks. This result is consistent with other studies showing that SMB is not priced in the cross section of portfolios sorted by size and book to market; see Adrian, Crump and Moench (2015) and Lettau and Livingston (2001).

Despite most of the factors are not statistically different from zero on average, hence not priced, they exhibit statistically significant time variation and fluctuate a lot between positive and negative values. Such time variation of the price of risk is well documented by the set of Figures 2.3 to 2.6. Figure 2.3 plots by columns the γs for the three different samples; with the top panel relating to 25 portfolios, the central panel to the 55 portfolios and bottom panel individual stocks. As before, the value of $h = 0.50$ and the $Polling$ methods describe a form of background path for the evolution of the price of risk while $Specific$ approach exhibits the highest volatility. From the analysis of these graphs is clear how much of the information about the price of risk is lost using approaches such $h = 0.50$, where we do not consider the specificity of each asset.

Table 2.4: Descriptive Statistics of risk premia estimates

	Obs.	Mean	St. Dev.	Min	Max	Skew	Kurt	SE	NW SE
25 Portfolio									
Rolling									
$\hat{\gamma}_0$	454	0.0086	0.0083	−0.0123	0.0371	0.4184	0.5990	0.0060	0.0054
$\hat{\gamma}_{\beta_{MRKT}}$	454	−0.0022	0.0088	−0.0226	0.0233	0.0766	−0.5119	0.0057	0.0050
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0021	0.0054	−0.0085	0.0145	0.4392	−0.7676	0.0010	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0032	0.0042	−0.0058	0.0168	0.2544	−0.2611	0.0011	0.0010
h=0.5									
$\hat{\gamma}_0$	454	0.0092	0.0082	−0.0067	0.0344	0.9116	1.3169	0.0059	0.0048
$\hat{\gamma}_{\beta_{MRKT}}$	454	−0.0024	0.0086	−0.0216	0.0129	−0.2629	−0.8884	0.0057	0.0046
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0042	−0.0045	0.0099	0.5480	−0.7747	0.0010	0.0009
$\hat{\gamma}_{\beta_{HML}}$	454	0.0028	0.0036	−0.0027	0.0099	0.2900	−1.0656	0.0011	0.0008
Polling									
$\hat{\gamma}_0$	454	0.0096	0.0074	−0.0044	0.0316	0.9550	1.0650	0.0059	0.0047
$\hat{\gamma}_{\beta_{MRKT}}$	454	−0.0028	0.0079	−0.0202	0.0103	−0.3081	−0.9680	0.0056	0.0045
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0037	−0.0043	0.0087	0.5381	−0.8492	0.0010	0.0009
$\hat{\gamma}_{\beta_{HML}}$	454	0.0028	0.0032	−0.0021	0.0086	0.1337	−1.1173	0.0011	0.0008
Average									
$\hat{\gamma}_0$	454	0.0098	0.0069	−0.0048	0.0350	0.8941	1.6039	0.0058	0.0046
$\hat{\gamma}_{\beta_{MRKT}}$	454	−0.0025	0.0073	−0.0216	0.0150	−0.1567	−0.6690	0.0056	0.0044
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0007	0.0037	−0.0113	0.0118	0.4682	0.2256	0.0009	0.0008
$\hat{\gamma}_{\beta_{HML}}$	454	0.0031	0.0032	−0.0069	0.0132	0.1595	−0.2499	0.0010	0.0008
Specific									
$\hat{\gamma}_0$	454	0.0064	0.0281	−0.1892	0.1286	−0.2239	5.9720	0.0161	0.0151
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0005	0.0275	−0.1089	0.1753	0.1478	4.6923	0.0155	0.0147
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0010	0.0061	−0.0158	0.0267	0.4351	1.6323	0.0036	0.0029
$\hat{\gamma}_{\beta_{HML}}$	454	0.0030	0.0086	−0.0284	0.0407	0.0598	2.1849	0.0039	0.0036
55 Portfolio									
Rolling									
$\hat{\gamma}_0$	454	0.0029	0.0060	−0.0083	0.0206	0.8131	0.4169	0.0032	0.0034
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0037	0.0074	−0.0137	0.0214	0.2648	−0.6650	0.0031	0.0034
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0019	0.0058	−0.0102	0.0146	0.3154	−0.8416	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0012	0.0052	−0.0127	0.0163	0.0905	0.1267	0.0012	0.0014
h=0.5									
$\hat{\gamma}_0$	454	0.0031	0.0060	−0.0080	0.0185	0.5166	−0.1746	0.0032	0.0031
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0039	0.0068	−0.0088	0.0177	0.3021	−0.7173	0.0031	0.0031
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0007	0.0047	−0.0055	0.0106	0.4547	−0.8907	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0004	0.0052	−0.0097	0.0091	0.0824	−0.8843	0.0012	0.0015
Polling									
$\hat{\gamma}_0$	454	0.0038	0.0052	−0.0049	0.0156	0.4892	−0.5169	0.0031	0.0029
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0032	0.0057	−0.0075	0.0142	0.2503	−0.8742	0.0030	0.0030
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0006	0.0038	−0.0049	0.0083	0.4662	−0.9538	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	0.0007	0.0043	−0.0090	0.0076	−0.2592	−0.3218	0.0012	0.0014
Average									
$\hat{\gamma}_0$	454	0.0043	0.0051	−0.0074	0.0179	0.2525	−0.4356	0.0030	0.0030
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0030	0.0057	−0.0107	0.0171	0.3331	−0.4746	0.0029	0.0030
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0004	0.0039	−0.0076	0.0097	0.2985	−0.7415	0.0010	0.0010
$\hat{\gamma}_{\beta_{HML}}$	454	0.0010	0.0044	−0.0098	0.0106	−0.2347	−0.2759	0.0012	0.0013
Specific									
$\hat{\gamma}_0$	454	0.0053	0.0109	−0.0385	0.0406	−0.0563	1.4520	0.0062	0.0063
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0020	0.0125	−0.0380	0.0517	−0.0727	1.9992	0.0061	0.0065
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0002	0.0066	−0.0220	0.0306	0.6872	1.7246	0.0026	0.0028
$\hat{\gamma}_{\beta_{HML}}$	454	0.0007	0.0076	−0.0228	0.0252	−0.0234	−0.0214	0.0028	0.0032
200 Stocks									
Rolling									
$\hat{\gamma}_0$	454	0.0013	0.0049	−0.0121	0.0110	−0.8412	0.5448	0.0011	0.0013
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0069	0.0095	−0.0198	0.0260	−0.6055	−0.3970	0.0024	0.0032
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0018	0.0062	−0.0125	0.0223	0.9650	0.6839	0.0013	0.0016
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0003	0.0062	−0.0164	0.0181	0.2282	−0.3997	0.0015	0.0017
h=0.5									
$\hat{\gamma}_0$	454	0.0023	0.0049	−0.0097	0.0114	−0.6056	0.5540	0.0011	0.0011
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0071	0.0101	−0.0117	0.0226	−0.4628	−1.2147	0.0024	0.0031
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0011	0.0045	−0.0056	0.0125	0.7757	0.0968	0.0013	0.0014
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0018	0.0062	−0.0140	0.0106	0.2524	−0.8254	0.0015	0.0017
Polling									
$\hat{\gamma}_0$	454	0.0026	0.0034	−0.0049	0.0085	−0.4529	−0.4068	0.0008	0.0008
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0068	0.0100	−0.0091	0.0179	−0.2393	−1.6461	0.0022	0.0025
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0024	0.0021	−0.0013	0.0083	0.4873	−0.0664	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0030	0.0023	−0.0062	0.0007	0.2737	−1.3960	0.0013	0.0015
Average									
$\hat{\gamma}_0$	454	0.0027	0.0034	−0.0070	0.0114	−0.2455	0.1043	0.0008	0.0008
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0065	0.0098	−0.0111	0.0182	−0.2348	−1.5546	0.0022	0.0025
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0023	0.0023	−0.0042	0.0107	0.3858	1.0965	0.0011	0.0011
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0028	0.0028	−0.0086	0.0082	1.0775	1.4751	0.0013	0.0015
Specific									
$\hat{\gamma}_0$	454.	0.0021	0.0049	−0.0145	0.0187	0.2124	0.1013	0.0001	0.0015
$\hat{\gamma}_{\beta_{MRKT}}$	454	0.0066	0.0157	−0.0493	0.0464	−0.5967	1.1184	0.0003	0.0042
$\hat{\gamma}_{\beta_{SMB}}$	454	0.0009	0.0080	−0.0240	0.0234	−0.2290	−0.0516	0.0001	0.0022
$\hat{\gamma}_{\beta_{HML}}$	454	−0.0004	0.0093	−0.0319	0.0304	−0.0736	0.2182	0.0002	0.0025

Note: Descriptive statistics of the estimated risk premia, computed for classical $FMcB$ approach and 4 different bandwidths specification: *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (as shown in Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset, Equation (2.8); *Specific*, multiple time varying bandwidths, one for each asset and time. The Newey West standard errors are also displayed in the last column. The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach for the computation of the time varying $RMSE$, as discussed in Equation (2.6).



(a) $\gamma 0$ (b) $\gamma MRKT$ (c) γSMB (d) γHML

Note: The figure provides the estimates of risk premia estimates computed using factor risk loadings calculated with Rolling window, with 5 years estimation period ($FMcB$ approach, green line) and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5* (black line); *Polling* a single value of h coming from the poll average the cross asset and time as shown in Equation (2.7)) (purple line); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation 2.8, blue line); *Specific*, multiple time varying bandwidths, one for each asset and time (red line). The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach for the computation of the time varying $RMSE$, as discussed in Equation (2.6).

Figure 2.3: Dynamic comparison of risk premia estimates for different approaches

In particular, only the $Specific$ approach seems able to capture the GFC, where the drop in the estimates of the price of the factors is clearly evident. In Figures 2.4 to 2.6 instead, we produce an analysis of the significance for the different estimates across time: Figure 2.4 contains the results for 25 portfolios, Figure 2.5 for 55 portfolios while Figure 2.6 the ones for individual stocks. All the Figures are structured as follow: in columns are reported the different γs while in each row there is a different method for the computation of the β as in Section 2. For what concerns the market risk premia, it shows a significant positive sign at the beginning of the sample until early 2000, when it becomes significantly negative. Such change has been captured by all the methods, despite it is more clear for the stock asset context.

TABLE 2.5: Percentage reduction of $RMSE$ for different model

	Bandwidth choice: $RMSE$		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}
25 Portfolio			
h =0.5	-0.522	-0.483	-2.302
Polled	-0.396	-0.168	-2.140
Average	-0.424	-0.444	-2.048
Specific	-5.800	-5.383	-6.738
55 Portfolio			
h =0.5	-0.528	-0.529	-2.235
Polled	-0.219	-0.196	-1.943
Average	-0.147	0.019	-1.871
Specific	-5.349	-4.444	-6.339
200 Stocks			
h =0.5	-0.535	-0.664	-1.902
Polled	0.033	-0.014	-1.382
Average	0.102	-0.107	-1.413
Specific	-2.393	-2.119	-3.310

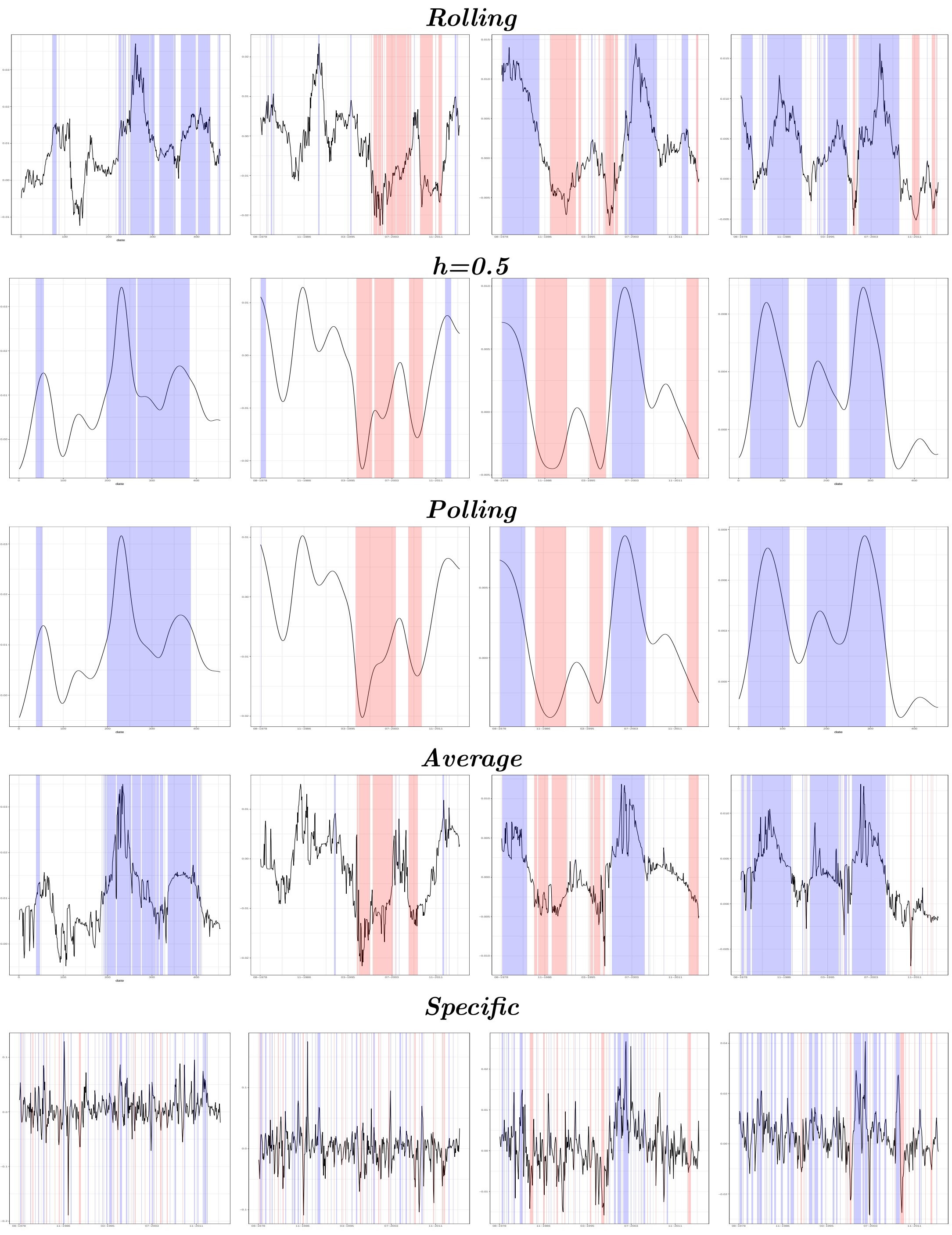
Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise as a percentage deviation from the benchmark model of Fama and MacBeth *Rolling*. The competing models are the followings: *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation (2.8)); *Specific*, multiple time varying bandwidths, one for each asset and time.

An important aspect of the our hierarchical method is the improvement in forecasting precision. Table 2.5 displays the $RMSE$ of an out of sample forecast exercise, reported as deviation from the $RMSE$ produced by the benchmark Fama and MacBeth (1973) approach (*Rolling*). The analysis of the tables identifies the $Specific$ approach as the best method, since it produces a remarkable reductions in the loss function and hence more precise forecast. The overall gains are greater for the portfolios of size 25 and 55. This approach produces improvements of around 6% with respect to the basic Fama and MacBeth (1973) and by 4.5% with respect to the kernel approach with optimal bandwidth parameter set to 0.5. Further, since the $Specific$ approach outperforms also the other two kernel methods, $Polling$ and $Average$, it is clear the importance of allowing the time variation in the bandwidth parameters, h, and optimising it for each asset. In line with Adrian, Crump and Moench (2015), we, further, observe that the classical rolling window approach is always outperformed by the kernel ones. Finally, we notice that in absolute terms increasing the sample size does not help to reduce the $RMSE$; the smallest values are reached performing the analysis for the 55 portfolio sample while the greatest for constituents of $S\&P500$.
Table 2.6 presents a pairwise analysis using the Diebold and Mariano (1995) test, henceforth DM, performed to certify the significance of the superior forecasting performance of the Gaussian kernel approach with time varying bandwidth. The *p-values* of the DM test are calculated under the null hypothesis that two competing models have the same predictive accuracy while the alternative is that the two method have significant different levels of accuracy. The analysis is conducted for all the sample and methods. The results are very striking and indicate that the DM test for the *Specific* method are statistically significant at the 0.01 level, confirming the aforementioned results. The key role of the time variation in the bandwidth parameter is also emphasized by the results of the method labelled $Average$ with respect to the $h = 0.5$ and $Polling$ approaches. Here, the null hypothesis of no difference in terms of performance cannot be rejected. In line with the literature, the Fama and MacBeth (1973) five year rolling window approach is never preferred to the kernel regression method with $h = 0.5.$, $Polling$ or $Average$. Ambiguous results instead are produced for the relations between $h = 0.5$ and $Polling$, where the former is preferred only in those cases where the $Polling$ has hs lower than 0.5, such 25 portfolio with $RMSE_{w=12}$. Such evidence is not unexpected given the nature and the characteristics of the $Polling$ approach.

Indeed, as Table 2.1 showed, increasing the sample size the degree of smoothness increases, producing flatter estimates that performs well in a forecasting exercise. Some further results on model comparisons and explanation of results are in Table 2.7; it displays the correlations between the beta estimates generated by each different asset. The $Specific$ approach is the method that produces less correlated estimates; the difference with respect the other methods is between 60% to 70%. This finding is robust to change in sample and methods for the choice of the bandwidth. Such results, with the fact that the $Specific$ approach provides small standard errors, let us solve two of the main critiques of the Fama and MacBeth (1973) (error in variable problem and cross sectional correlation), remaining agnostic on the choice of data between portfolio and individual stock; see Shanken (1992) Adrian and Franzoni (2009) among the others.
These findings extend the results of Adrian *et al.* (2015) and are consistent with those of Ferson and Harvey (1991), highlighting the importance of using not only a dynamic framework but also dynamic estimation approach with minimal theoretical restriction.

2.6.1 Robustness checks

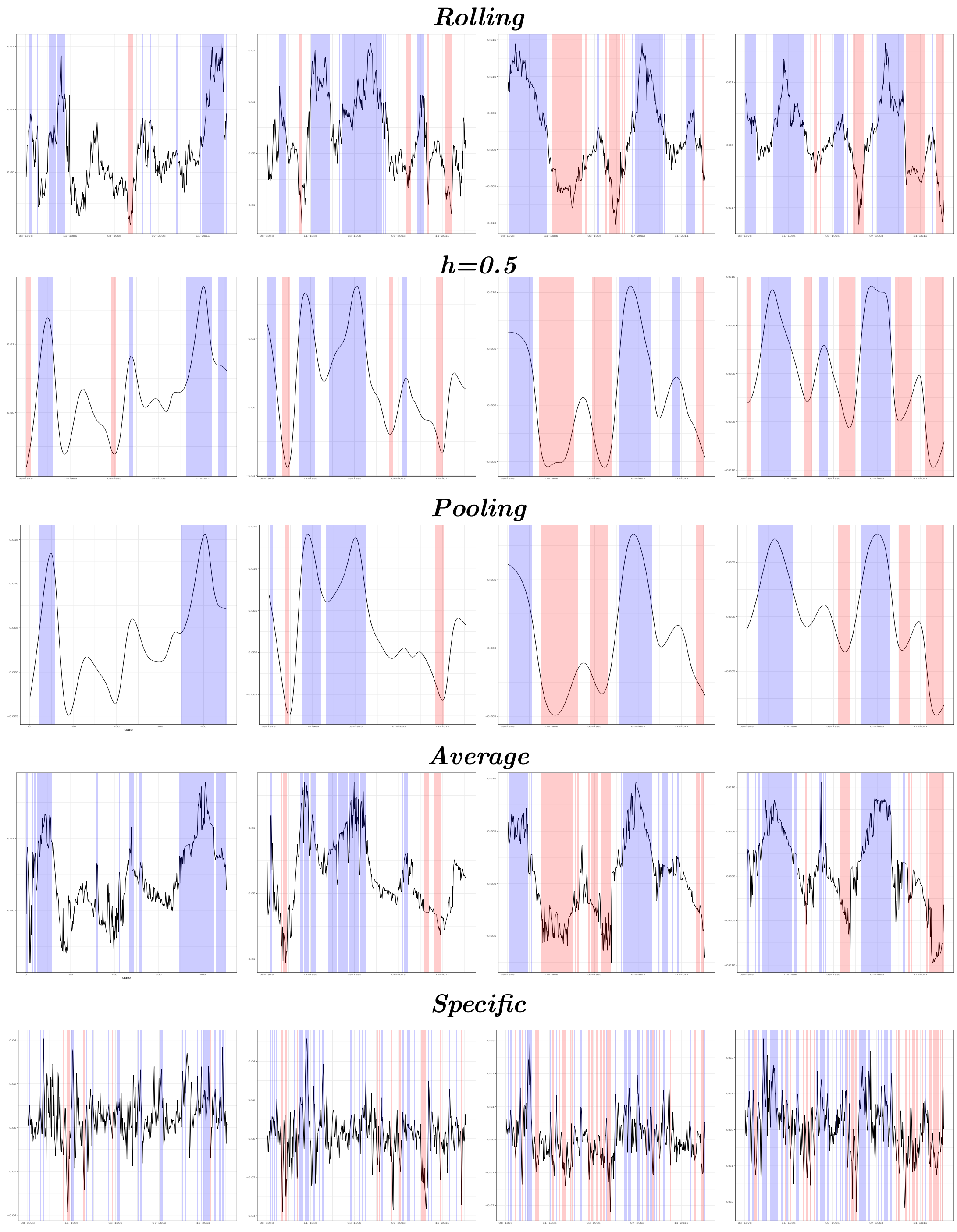
A substantial number of robustness checks were performed to test the aforementioned findings. Full details are available in Appendix F, where we report the $RMSE$ for each approach in terms of deviation from the benchmark $FMcB$.
Firstly, we investigate the sensitivity of the results to the choice of bandwidth parameter range orignally set as $[0.05; 0.9]$. We analyse three alternative intervals: $[0.35; 0.9]$, $[0.05; 0.6]$ and $[0.25; 0.75]$. The results, reported in Table F.1, confirm that the $Specific$ approach leads to a reduction of $RMSE$ that oscillates between 0.6% and 8.2% for the 25 and 55 Portfolio and between 0.5% and 6.4% for constituents of $S\&P$ 500.



(a) γ_0 (b) γ_{MRKT} (c) γ_{SMB} (d) γ_{HML}

Note: The figure provides significance analysis of the estimates of risk premia estimates. The blue areas are periods in which the estimates are statistically positive at 5% level of significance, while the red ones identify period in which the estimates are negative. The series have been computed using different approaches: Rolling window ($FMcB$ approach), *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation (2.8)); *Specific*, multiple time varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach as discussed in Equation (2.6).

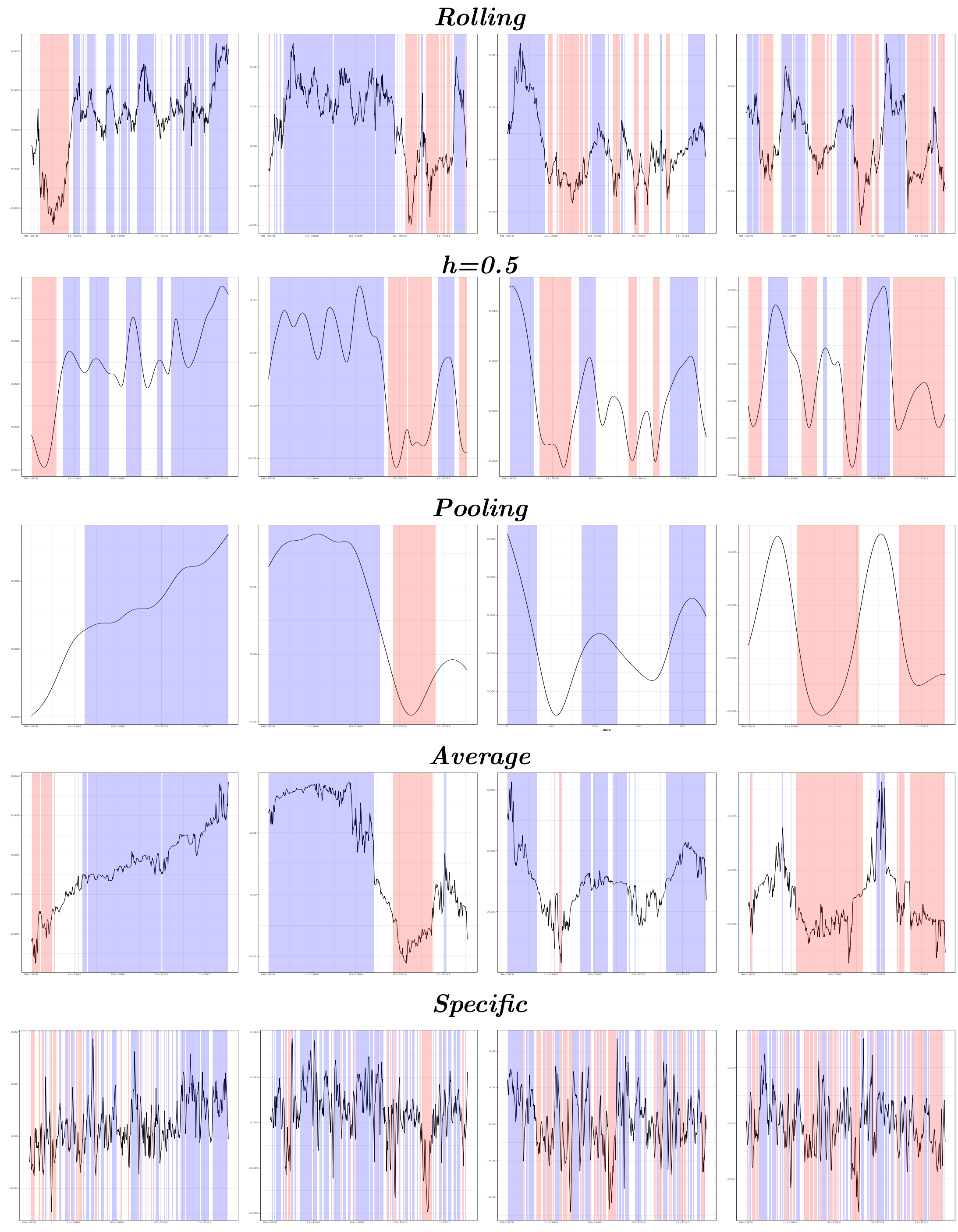
FIGURE 2.4: Comparison of γs significance of different approaches - 25 Portfolios



(a) γ_0 (b) γ_{MRKT} (c) γ_{SMB} (d) γ_{HML}

Note: The figure provides significance analysis of the estimates of risk premia estimates. The blue areas are periods in which the estimates are statistically positive at 5% level of significance, while the red ones identify period in which the estimates are negative. The series have been computed using different approaches: Rolling window ($FMcB$ approach), *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation (2.8)); *Specific*, multiple time varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach as discussed in Equation (2.6).

FIGURE 2.5: Comparison of γs significance of different approaches - 55 Portfolios



(a) γ_0 (b) γ_{MRKT} (c) γ_{SMB} (d) γ_{HML}

Note: The figure provides significance analysis of the estimates of risk premia estimates. The blue areas are periods in which the estimates are statistically positive at 5% level of significance, while the red ones identify period in which the estimates are negative. The series have been computed using different approaches: Rolling window ($FMcB$ approach), *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation (2.8)); *Specific*, multiple time varying bandwidths, one for each asset and time. The choice of the optimal bandwidth parameter, h_t^{opt}, has been made using the kernel approach as discussed in Equation (2.6).

FIGURE 2.6: Comparison of γs significance of different approaches - 200 stocks

Table 2.6: Diebold and Mariano test results

	25 Portfolio				***55 Portfolio***				***200 Stocks***			
	Rolling	**h**$= 0.5$	**Polling**	**Average**	**Rolling**	**h**$= 0.5$	**Polling**	**Average**	**Rolling**	**h**$= 0.5$	**Polling**	**Average**
$h_{w=12}$												
h $= 0.5$	0.0179				0.0144				0.0193			
Polling	0.2688	0.0407			0.4797	0.0426			0.9222	0.1062		
Average	0.2401	0.4732	0.8207		0.6468	0.0767	0.4621		0.7755	0.0235	0.8493	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0018	0.0061	0.0043	0.0005
$h_{w=24}$												
h $= 0.5$	0.0202				0.0134				0.0113			
Polling	0.5511	0.0569			0.3998	0.0542			0.9527	0.1051		
Average	0.1953	0.7605	0.0466		0.9368	0.0392	0.1147		0.7303	0.0183	0.7439	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0044	0.0025	0.0007
h_{Kern}												
h $= 0.5$	0.0000				0.0001				0.0007			
Polling	0.0001	0.0521			0.0001	0.0542			0.0088	0.1181		
Average	0.0001	0.2588	0.5973		0.0004	0.1260	0.5895		0.0098	0.1721	0.5460	
Specific	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: The table provides the p values of the DM test applied on the results of the Table 2.4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy, while the alternative is that the two methods have significant different level of accuracy for the out of sample one step ahead forecasting exercise.

Table 2.7: Correlation matrix among factor risk loadings

	25 Portfolio			*55 Portfolio*			*200 Stocks*		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
β_{MRKT}									
Rolling	0.2969	0.3216	0.2747	0.3118	0.3359	0.2888	0.3074	0.3183	0.2936
h=0.5	0.3159	0.3185	0.3141	0.3334	0.3381	0.3317	0.3293	0.3335	0.3301
Pooling	0.3284	0.3608	0.3303	0.3785	0.4235	0.3782	0.5030	0.5966	0.4945
Average	0.3106	0.3389	0.2894	0.3663	0.4114	0.3491	0.4525	0.5235	0.4451
Specific	0.0887	0.0937	0.0892	0.0985	0.1108	0.0967	0.1026	0.1220	0.0755
β_{SMB}									
Rolling	0.3155	0.3379	0.2995	0.2984	0.3204	0.2797	0.4223	0.4314	0.4187
h=0.5	0.3335	0.3356	0.3267	0.3266	0.3257	0.3293	0.5252	0.5368	0.4829
Pooling	0.3491	0.3864	0.3471	0.3721	0.4076	0.3740	0.6785	0.7183	0.6254
Average	0.3406	0.3774	0.3165	0.3549	0.3838	0.3403	0.6179	0.6618	0.5942
Specific	0.0940	0.0930	0.0884	0.0967	0.0978	0.0900	0.1546	0.1895	0.1151
β_{HML}									
Rolling	0.4920	0.5310	0.4525	0.4417	0.4831	0.4007	0.3418	0.3729	0.3160
h=0.5	0.5263	0.5310	0.5214	0.4689	0.4744	0.4643	0.3625	0.3624	0.3575
Pooling	0.5467	0.5808	0.5488	0.5219	0.5650	0.4744	0.5373	0.6080	0.5309
Average	0.5082	0.5576	0.4826	0.4917	0.5335	0.4702	0.4952	0.5339	0.4774
Specific	0.1150	0.1135	0.1065	0.1094	0.1167	0.0998	0.1024	0.1271	0.0838

Note: The table provides the average correlation among the 3 factor loadings for all the approaches under analysis: Rolling window, with 5 years estimation period and kernel weighted regressions using 4 different optimal bandwidth; *h=0.5*; *Polling* a single value of h coming from the poll average the cross asset and time (Equation (2.7)); *Average*, a unique time varying bandwidth coming from the average of h across asset (Equation (2.8)); *Specific*, multiple time varying bandwidths, one for each asset and time, h^{opt}.

Further, the awareness of possible over fitting issues due to the combination of sample size of the training period and the small value for the bandwidth parameters, lead us to investigate also the specification of the bandwidth parameter using a time varying $LASSO$ approach inside our hierarchical methodology. After an accurate analysis for the choice of the penalization parameter, we decide to use values of λ that allow us to maintain the model unchanged[4] ($\lambda \in [0.00005; 0.000001]$). The results, displayed in Table F.3, shows that increasing the penalisation we significantly increase the gain of our $Specific$ technique, now between 2.8% to 10.2%.

To avoid possible presence of over fitting concerns, Table F.4 shows how changes in the size of training period affects the results. We investigate results using ten years of data, $T = 120$ observations and fifteen years, $T = 180$ observations. Table F.4 confirms

[4]The purpose of the paper is not to identify the best factors but the identification of the best methodology. Therefore, to guarantee an identical setting, we maintain the factors in the model fixed.

that the *Specific* outperforms the benchmark model and is the one with the highest reduction of the $RMSE$. In Table F.2, we perform an analysis changing the sample period in order to exclude the global financial crisis. The new sample tested is 1973-2007. The results confirm again that *Specific* approach outperforms its competitors. Finally, since the goal of this paper is to propose a new estimation method to increase the forecasting performances of any asset pricing model, we consider different model specification: momentum factor by Carhart (1997) and the 5 factor model by Fama and French (2015). Once again the results in Table F.5 confirms the robustness of our findings: *Specific* delivers once again the lowest reduction in the loss function.

2.7 Concluding remarks

Producing accurate estimates of risk premia is a key feature in many financial activities, including asset pricing, corporate finance and risk management. From a pricing perspective, during last decades a plethora of works has shown the existence of time variation in the risk exposures, and proposed several different dynamic asset pricing models to capture it, see Jagannathan and Wang (1996), Ferson and Harvey (1999), and Lettau and Ludvigson (2001) Ang and Kristensen (2012) among the others.
This paper has developed a new framework for the estimation of beta coefficients for a generic dynamic asset pricing model that imposes little a priori structure and generalizes the classic two step Fama and MacBeth (1973) procedure. Time variation in the beta estimates is found from a kernel weighted regression that significantly improves on conventional results in a $RMSE$ sense. We use a cross validation procedure which allows us to optimise the choice of the time bandwidth parameter for each asset at each point in time. This very flexible approach, without imposing an extensive a priori structure, improves the estimation of the risk premia. The empirical results overwhelmingly show that time variation of risk associated with stocks and portfolios must be captured with an estimation procedure that on one hand avoids imposing excessive a priori structure and on the other hand, takes into account the specific features of each asset and the time variation of its generating mechanism. Our methodology is indeed able to produce an increase in the forecasting performance greater (between 4% to 7%) than the alternative methods and independently for any type of model and asset.

Chapter 3

Long Memory, Realized Volatility and Heterogenous AutoRegressive Models

3.1 Summary

The presence of long memory in Realized Volatility (RV) process is a widespread stylized fact well documented in financial literature. The origins of long memory in RV have been attributed to jumps, structural breaks, contemporaneous aggregation, non-linearities, or pure long memory. An important development in modelling the RV has been the Heterogeneous Autoregressive (HAR) model and its extensions. This paper assesses the separate roles of fractionally integrated long memory models, extended HAR models and time varying parameter HAR models in modelling the RV. We find that the presence of the long memory parameter is often important in addition to the HAR models.

3.2 Introduction

The long memory feature of many time series has long been of interest to statisticians, econometricians and researchers in many of the physical sciences, who have become aware of the very strong persistence in the autocorrelations and other measures of the temporal dependence of some time series. Hurst (1951, 1957) and Mandelbrot and Wallis (1968) noted the phenomena of long memory in river flow and hydrological data; and Greene and Fielitz (1977) in financial data. Some of the historical developments are discussed by Baillie (1996). One of the fascinations with long memory processes is their inherent ability to bridge both persistent stationary and non stationary time series. One of the most ubiquitous and also important examples of long memory are to be found in application to Realized Volatility (RV) time series.

The construction of observable RV series from high frequency financial market data has now become standard practice in empirical finance. One of the attractions with using RV is to reduce emphasis of the formulation and choice of model, with a direct measurement of volatility. It has been found that RV time series are characterized by very strong persistence in their autocorrelations for a wide range of financial assets. An interesting issue has been to provide an explanation for this phenomenon and to assess whether it could be due to jumps, structural breaks, omitted non-linearities, contemporaneous aggregation, or to just "pure long memory".

However, a popular way of describing RV has been the Heterogeneous Autoregressive (HAR) model, which was originally due to Corsi (2009). The model is based on an additive cascade of partial volatilities from high frequencies to low frequencies with each additive cascade having close to an $AR(1)$ structure. This idea of multiple components in the volatility process has been justified in terms of the differences of agents risk profiles, institutional structures, temporal horizons, etc. In general, the HAR model appears attractive as a simplified regression based procedure for approximating the persistence of many RV time series.

This paper examines the relationship between long memory models, the HAR model and the extended versions of the HAR model, which include semi variances, signed jump variations, and "good" and "bad" volatility. We estimate HAR models from simulated fractional white noise processes and find the simulated estimates have certain similarities with the HAR estimates from actual RV data. We also estimate by MLE an $ARFIMA$ model which includes the regular long memory parameter plus an autoregressive component of order 22, which embodies the parameter restrictions implied by the HAR model. The model is essentially a restricted $ARFIMA$ and is denoted as an $RARFIMA(22, d, 0)$ model. The model is theoretically quite similar to the basic HAR model with long memory disturbances and is also estimated by an MLE procedure. The overall conclusion is that in many cases both the long memory feature and the HAR structure for short and medium term memory can be important in representing variation within RV series.

Finally, we also consider a time varying parameter, kernel weighted regression approach to estimate HAR models. These estimated models indicate that the relative importance of the partial volatility cascades typically varies throughout the samples. Such a Time Varying Parameter (TVP) model, denoted by $TVP - HAR$, is quite effective in representing some of the long memory characteristics of RV time series. However, model selection information based criteria generally favour the simpler $RARFIMA$ structure with constant long memory and HAR parameters.

The plan of the rest of the paper is as follows: Section 3.3 defines some of the theoretical aspects of RV and also includes details of the statistical quantities regularly implemented and arising from RV series. Section 3.4 briefly describes the RV data and some of their basic characteristics; while Section 3.5 describes the long memory models and inferential methods and report MLE of both $ARFIMA$ models and also reports semi parametric estimation of the long memory parameter. Section 3.6 describes the various HAR models and their estimates, including various extensions including jumps and good and bad volatility components. Section 3.7 is concerned with different methods for attempting to distinguish between HAR and long memory and also for combining these approaches. In particular, we provide simulation evidence on the properties of OLS estimation of HAR models when the true data generating process is a fractional white noise, long memory process. This section also includes results on the MLE of unrestricted $ARFIMA(22, d, 0)$ and $RARFIMA(22, d, 0)$ models, where the restrictions are from the HAR formulation. We also include MLE of extended HAR models which have long memory disturbances. Section 3.8 describes an alternative approach based on a time varying parameter HAR model which involves kernel weighted regressions with time varying regression coefficients based on the method by Giraitis *et al.* (2014). Section 3.9 discusses some of the results concerning comparisons of the models and also provides a brief conclusion.

3.3 Basics of Realized Volatility

The variable RV is a model free measurement of financial market volatility and was proposed by Andersen *et al.* (2001, 2003) and Barndorff-Nielsen and Shephard (2002). We define a continuous time diffusion process for the log of price (p_t) as

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \;\; t \geq 0,$$

where $dp(t)$ is the change in the logarithmic price, $\mu(t)$ denotes the drift term which has continuous and locally bounded variations, $\sigma(t)$ is a strictly positive volatility process and $W(t)$ is standard Brownian motion. Assuming a unit for the time length of one day, daily returns can be expressed as

$$r_t = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s).$$

The volatility of an asset's returns is related to the evolution of the spot volatility (σ_t) so that the distribution of returns depends on both the drift and spot volatility components; hence

$$r_t \sim N\left(\int_{t-1}^{t} \mu(s)ds, \int_{t-1}^{t} \sigma^2(s)\right).$$

RV at day t is RV_t and is defined as the sum of high frequency, intraday squared returns. Hence

$$RV_t = \sum_{\tau=1}^{m} r_{t,\tau}^2,$$

where $r_{t,\tau} = p_{t,\tau} - p_{t,\tau-1}$ is the intraday return based on m intraday log-prices of the asset $\{p_{t,\tau}\}_{\tau=1}^{m}$ within day t observed at m fixed time intervals of $\tau = 1, \ldots, m$. Andersen *et al.* (2003) showed that under suitable conditions, including the absence of serial correlation in the intraday returns, RV_t is a consistent estimator of integrated volatility (IV_t). Hence

$$RV_t = \sum_{\tau=1}^{m} r_{t,\tau}^2 \xrightarrow{p} \int_{t-1}^{t} \sigma_s^2 ds.$$

The basic RV model has been extended to include the effects of jump components. Suppose the log-price process is a Brownian Semi-Martingale with Jumps, then

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t) \quad t \geq 0,$$

where the jump component is $\kappa(t)dq(t)$, with $\kappa(t)$ as the size of the jump and $dq(t)$ as a continuous process with $dq(t) = 1$ if there is a jump at time t and is 0 otherwise. The corresponding discrete-time daily returns are

$$r_t = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma^2(s)dW(s) + \sum_{j=N(t-1)+1}^{N(t)} \kappa(s_j),$$

where $N(t)$ counts the number of jumps occurring with possibly time varying intensity and jump size $\kappa(s_j)$ and where s_j are the jump times. In the presence of jumps, RV_t converges uniformly in probability to

$$RV_t \xrightarrow{p} \int_{t-1}^{t} \sigma^2\left(s\right)ds + \sum_{j=N(t-1)+1}^{N(t)} \kappa^2(s_j).$$

Hence, RV_t is a consistent estimator of IV_t only in the absence of jumps, while otherwise it converges to a quantity that also accounts for the jump process, $\sum_{j=N(t-1)+1}^{N(t)} \kappa^2(s_j)$. Hence RV provides an ex-post measure of the true total variation including the discontinuous jump part. The above assumes independence of the quantities $W(s)$, $N(t)$ and $\kappa(s_j)$.

To decompose volatility into a component that relates only to positive high-frequency returns and a component that relates only to negative high-frequency returns, we use the realized semi variance quantity proposed by Barndorff-Nielsen and Shephard (2007). The positive (negative) realized semi variance RS_t^+ (RS_t^-) is computed by summing the squared intra-day returns associated with an increase (decrease) in the asset price. Then,

$$\begin{aligned} RS_t^+ &= \sum_{\tau=1}^{m} r_{t,\tau}^2 I\left\{r_{t,\tau} > 0\right\} \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} \Delta p_s^2 I\left\{\Delta p_s > 0\right\}, \\ RS_t^- &= \sum_{\tau=1}^{m} r_{t,\tau}^2 I\left\{r_{t,\tau} < 0\right\} \xrightarrow{p} \frac{1}{2}\int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1<s\leq t} \Delta p_s^2 I\left\{\Delta p_s < 0\right\}, \end{aligned}$$

where $I\left(\cdot\right)$ is an indicator function and $\Delta p_s = p_s - p_{s-}$ captures a jump, if present. The notation above is used to be consistent with Patton and Sheppard (2015) and it should be noted that some less cumbersome schemes of notation can be used which involve using $\Delta p_j = p(s_j) - p(s_{j-})$ and summations over j. Also, note that $RV_t = RS_t^+ + RS_t^-$.

Following Patton and Sheppard (2015), we compute the signed jump variation as

$$\Delta J_t^2 = \left(RS^+ - RS^-\right) \xrightarrow{p} \sum_{t-1<s\leq t} \Delta p_s^2 I\left\{\Delta p_s > 0\right\} - \sum_{t-1<s\leq t} \Delta p_s^2 I\left\{\Delta p_s < 0\right\}.$$

Note that the continuous part of RV cancels out and only the jump components remain. We analyse whether the impact of jumps depends on the sign of positive and negative jump variation. Hence, following Patton and Sheppard (2015), we further decompose the signed jump variation as

$$\begin{aligned} \Delta J_t^2 &= \Delta J_t^{2+} + \Delta J_t^{2-} \qquad (3.1) \\ &= \left(RS_t^+ - RS_t^-\right) I\left\{RS_t^+ - RS_t^- > 0\right\} + \left(RS_t^+ - RS_t^-\right) I\left\{RS_t^+ - RS_t^- < 0\right\}. \end{aligned}$$

Consistent estimation of the continuous part of the volatility, or IV, has been achieved by Barndorff-Nielsen and Shephard (2004), who proved that under the regularity condition that jumps have finite activity, the normalized sum of products of the adjacent absolute values of returns, i.e. Bipower Variation (BV), is a consistent estimator of IV even in the presence of jumps. At day t, BV is defined as

$$BV_t^0 = \frac{\pi}{2} \sum_{\tau=2}^{m} \left|r_{t,\tau}\right| \left|r_{t,\tau-1}\right| \underset{p}{\rightarrow} \int_{t-1}^{t} \sigma_s^2 ds \qquad \text{as } m \rightarrow \infty.$$

The above critically assumes independence of the previously defined quantities $W(s)$, $\sigma^2(s)$, $N(t)$ and $\kappa(s_j)$. Rather than using BV^0 directly, we use an average of skip-0 through skip-4 BV estimators as in Patton and Sheppard (2015),

$$BV_t = \frac{1}{5} \sum_{q=0}^{4} BV_t^q,$$

where skip-q BV estimator is defined as

$$BV_t^q = \frac{\pi}{2} \sum_{\tau=q+2}^{m} \left|r_{t,\tau}\right| \left|r_{t,\tau-1-q}\right|.$$

The skip-q BV estimator corrects small sample bias of the skip-0 BV estimator.
Over the last few years, many techniques have been proposed to estimate, or to at least proxy, asset return volatility from high frequency data. See Meddahi *et al.* (2011) and

Andersen *et al.* (2006) for details. Some methods have focused on correcting for microstructure noise caused by trade imperfections, market frictions, or informational effects. The most commonly used technique for computing RV is known as *down-sampling*, which conventionally uses sampling intervals from 5 to 30 minutes to derive daily RV series. This method does not use all the high frequency data; and other methods have been suggested in the literature to try to deal with the presence of possible micro-structure noise. In particular, Bandi and Russell (2008) have considered the idea of finding the optimal sampling frequency; while Aït-Sahalia *et al.* (2005) have used an MLE of a model for RV which assumes additive *i.i.d.* microstructure noise. Zhou (1996) considered corrections for first-order autocorrelation type noise in high frequency data; Barndorff-Nielsen *et al.* (2008) use a realized kernel to correct for autocorrelation in a more general approach. Zhang *et al.* (2005) and Zhang (2006) use two-scale and multi-scale estimators which combine sub sampled RV computed at lower and higher frequencies.

However, as noted by Liu *et al.* (2015), there is uncertainty as to the desirability and also choice of the most appropriate method. After considerable amount of initial data analysis and investigation of possible outliers and noise, we decided to use 5-minute data for the computation of RV. This seemed the most appropriate method for calculating RV given the purpose of this study is to compare, contrast and to combine long memory and the HAR modelling approaches.

3.4 Data

In order to assess the relative merits of HAR and long memory models we use five minute high-frequency, intraday returns data on various assets. We examine five spot exchange rates of the Australian dollar (AUD), the Canadian dollar (CAD), the Euro (EUR), the UK British pound (GBP), and the Japanese yen (JPY) all against the numeraire US dollar (USD); for the period, January 2, 2004 through December 29, 2017. In line with previous studies we exclude the slower trading patterns induced over the weekends by discarding all observations from Friday 21:00 GMT through Sunday 22:00 GMT and measure the rates as the midpoint of the logarithms of the bid and ask rates. This provides a sample size of $T = 3,627$ daily observations from which to compute RV and the semi-variance measures.

For the equity market data, we use the $S\&P500$ index which consists of five minute tick interpolated prices from January 2, 2001 through December 31, 2016. The trading hours span from 9:30 through 16:00 with a total of 78 intra-day observations and the total number of the trading days after adjustments is $T = 4,172$ observations.

Figure 3.1 plots the time paths of the various RV series and Figure 3.2 plots the first 50 lags of the sample autocorrelation function of the RV series. It can be seen that all the autocorrelation functions for the RV series exhibit the strong persistence that is consistent with long memory behaviour.

3.5 Long memory and Realized Volatility

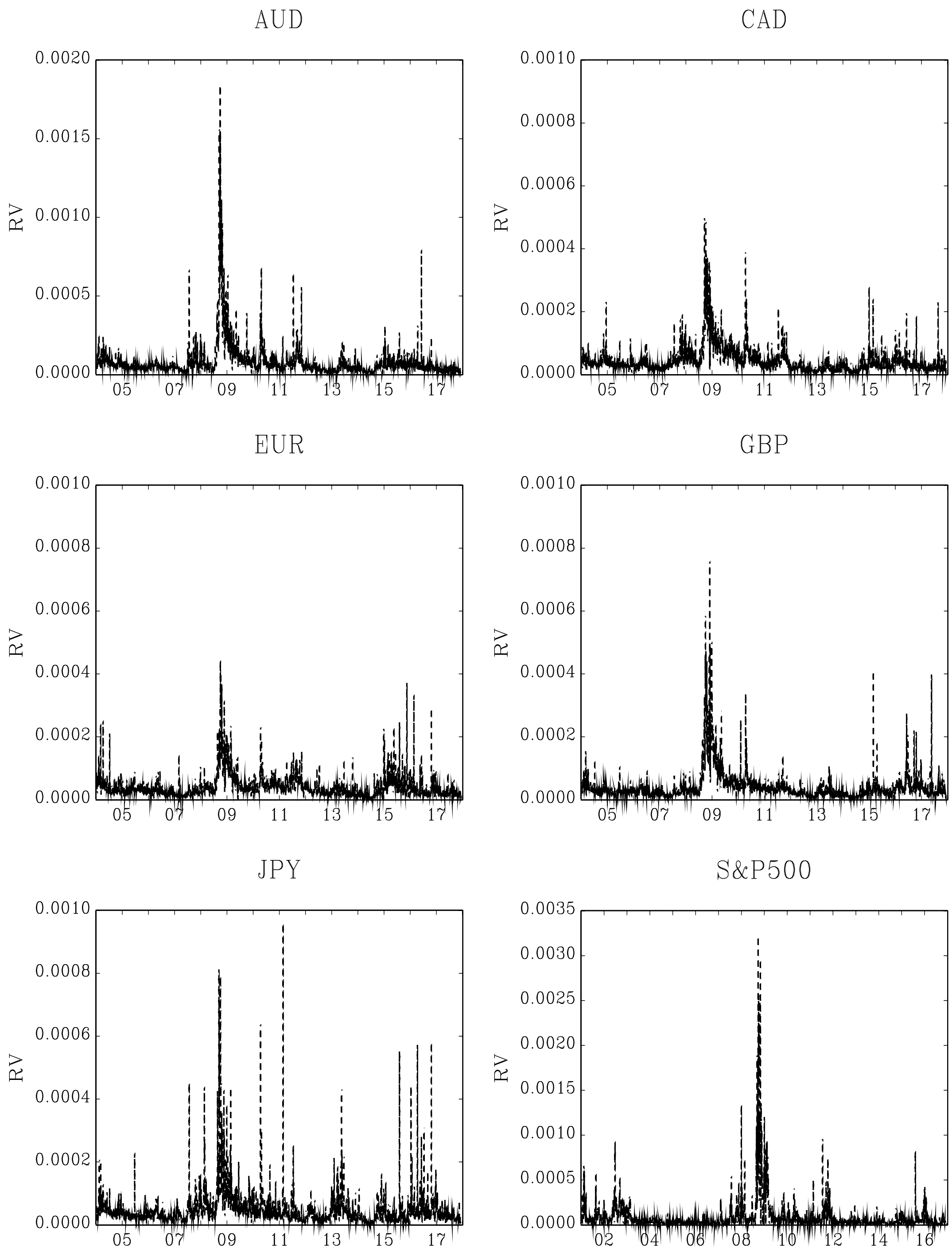
Some of the statistical features of the various RV series may be described in terms of the fractionally integrated, or long memory time series process, as defined by

$$(1-L)^d y_t = u_t,\ t = 1, \ldots, T,$$

where L is the lag operator, u_t is a short memory, $I(0)$ process, and the observable time series y_t is defined to be fractionally integrated of order d, or $I(d)$. In this case y_t is generally the RV series. The process generates hyperbolic rates of decay in the autocorrelation function and Impulse Response Function (IRF). The $I(d)$ process is defined as having partial sums that converge weakly to fractional Brownian motion, while d represents the degree of "long memory", or persistence in the series. For $-0.5 < d < 0.5$ the process is stationary and invertible, while for $0.5 \leq d \leq 1$, the process does not have a finite variance. However, the IRF does decline to zero for $d < 1$. The IRF, or infinite order moving average representation of this process, is given by

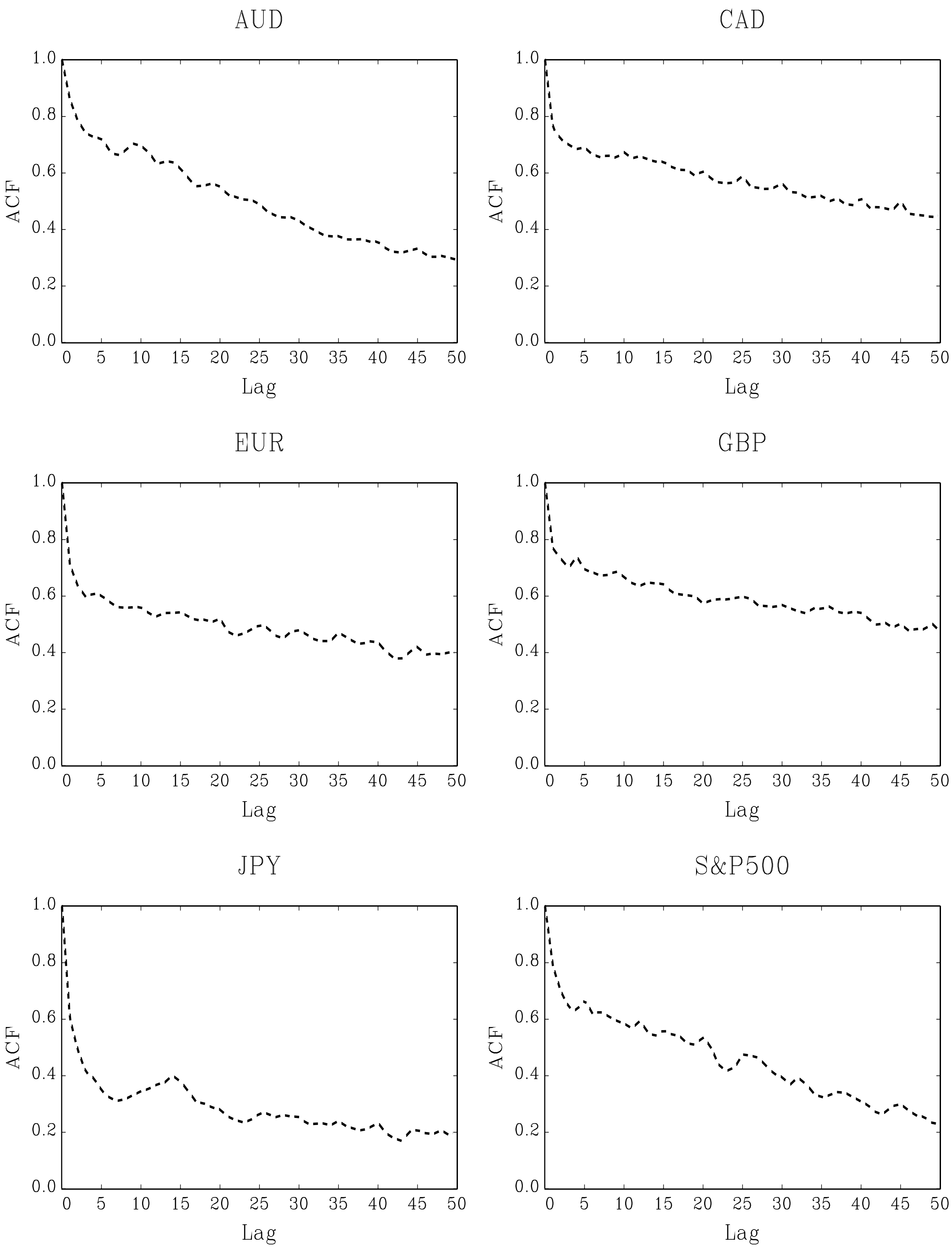
$$y_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k},$$

where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E(\epsilon_t \epsilon_s) = 0$, $s \neq t$. For large lags k, these coefficients decay at the very slow hyperbolic rates of $\psi_k \sim c_1 k^{d-1}$ and similarly the infinite autoregressive representation coefficients decay at the rate of $c_2 k^{-d-1}$ and autocorrelation coefficients at the rate of $c_3 k^{2d-1}$, where c_1, c_2 and c_3 are constants. The



Note: The figure provides the plots for Realized Volatility for different exchange rates, computed as shown in Equation (3.3).

FIGURE 3.1: Realized Volatility for each financial series



Note: The figure provides the autocorrelation functions for Realized Volatility for different exchange rates, computed as shown in Equation (3.3).

Figure 3.2: Autocorrelation functions for the *RV* series of financial series

simplest discrete time parametrization is the $ARFIMA$ model, which combines long memory with short run $I(0)$ dynamics and provides a flexible extension of the $ARIMA$ model and was introduced by Granger (1980), Granger and Joyeux (1980), and Hosking (1981). The simplest time domain workhorse model for long memory processes is the $ARFIMA(p, d, q)$ model of the form

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t, \tag{3.2}$$

where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator of orders p and q respectively. Maximization of the Gaussian log likelihood is accomplished with respect to the complete vector of parameters $\vartheta' = \left(d, \phi_1, ...\phi_p, \theta_1, ...\theta_q, \sigma^2\right)$. Under these conditions, the asymptotic distribution of the MLE will be

$$T^{1/2}\left(\hat{\vartheta} - \vartheta_0\right) \to N\{\mathbf{0}, \mathbf{I}(\vartheta_0)^{-1}\},$$

where ϑ_0 denotes the true value of the vector of parameters and $\mathbf{I}(\vartheta_0)$ is the information matrix. The results follow from Fox and Taqqu (1986) and for sake of simplicity a demeaned process is assumed. Then the MLE are $T^{1/2}$ consistent and asymptotically Normal when the unconditional mean is zero or known. The inclusion of an intercept parameter will result in a $T^{1/2-d}$ consistent estimator of the intercept. In some circumstances the assumption of Gaussianity may be inappropriate and can be replaced with the assumption that the innovations in Equation (3.2) merely satisfy some mild mixing conditions. Given the results in Hosoya (1997), the implementation of quasi MLE is then straightforward; and in particular,

$$T^{1/2}\left(\hat{\vartheta} - \vartheta_0\right) \to N\{\mathbf{0}, \mathbf{A}(\vartheta_0)^{-1}\mathbf{B}(\vartheta_0)\mathbf{A}(\vartheta_0)^{-1}\},$$

where $\mathbf{A}(\cdot)$ is the Hessian and $\mathbf{B}(\cdot)$ is the outer product gradient, both of which are evaluated at the true parameter values ϑ_0; see Baillie and Kapetanios (2008, 2013) for further details.

It is worth noting that long memory characteristics can be induced in a time series by many mechanisms. In particular, Granger (1980) showed that the aggregation of contemporaneous stationary $AR(1)$ processes could lead to an aggregate process with fractional integration. Also, occasional break points as in Granger and Hyung (2004);

and forms of regime switches, as shown by Diebold and Inoue (2001), can also give rise to the appearance of long memory. In many instances there may not be any obvious explanation as to the occurrence of long memory in time series data. However, such fractional processes can simply be regarded as more general forms of the Wold decomposition than the exponential decay implied by processes with rational spectra, or stationary and invertible $ARMA$ representations. Hence, in some sense, hyperbolic rates of decay do not appear any more in need of justification than the standard exponential rates of decay.

Table 3.1 reports the MLE of $ARFIMA(p, d, 0)$ models where the order p is selected on the basis of minimizing Schwarz (1978) BIC for $p \in \{0, 1, 2, ..., 10\}$. This is predicated on the assumption that the short memory components are sufficiently well approximated by a finite order $AR(p)$ process. For some of the RV series it was found necessary to have quite high order autoregressive components to deal with fairly substantial short memory $I(0)$ components in addition to the long memory property. The estimated long memory parameters were statistically significantly different from zero for all of the RV series with several cases of borderline non stationarity, which still imply finite cumulative $IRFs$. In all cases the estimates of the short memory parameters are suppressed in the interests of conserving space and all the emphasis is on the estimation of the long memory parameter, d. A more parsimonious parametrization of the short memory component can theoretically be found from the $ARFIMA(p, d, q)$ model; and estimates of this model are also reported in Table 3.1. The strategy for model selection of $ARFIMA(p, d, q)$ requires estimation of models of orders of $p \in \{0, 1, 2, ..., P\}$ and $q \in \{0, 1, 2, ..., Q\}$ where P and Q are the maximum orders of the short memory parameters being considered. In this study $P = Q = 8$, so that the implementation of minimizing the BIC required estimation of 81 models. It should be noted that while the $ARFIMA(p, d, q)$ models are expected to provide a more parsimonious parametrization of the short memory components, their use can be complicated due to near cancellation of AR and MA roots. In general there is reasonable consistency across the time domain results with borderline non stationary fractional integration for many of the RV series; and we conclude that the RV series appears to be quite well suited to be represented by the fractionally integrated $ARFIMA$ models.

Following the seminal paper of Geweke and Porter-Hudak (1983), an alternative procedure is to use semi parametric estimation of the long memory parameter, which

TABLE 3.1: Estimates of Long Memory Parameter d

	AUD	CAD	EUR	GBP	JPY	S&P500
			$ARFIMA(p,d,0)$			
p	7	4	3	4	0	6
d	0.786	0.625	0.553	0.598	0.396	0.728
	(0.114)	(0.067)	(0.051)	(0.099)	(0.045)	(0.180)
$ln(L)$	–2919.626	14.173	84.690	–714.267	–2106.755	–6086.334
BIC	5921.213	29.028	–120.202	1485.907	4238.098	12247.694
			$ARFIMA(p,d,q)$			
p	7	3	3	3	2	3
q	2	3	3	5	3	1
d	0.725	0.793	0.670	0.685	0.488	0.733
	(0.140)	(0.210)	(0.132)	(0.193)	(0.104)	(0.202)
$ln(L)$	–2875.159	83.275	123.603	–676.273	–2073.290	–6170.745
BIC	5848.672	–92.785	–173.440	1442.704	4212.150	12399.843

Note: The $ARFIMA(p,d,0)$ models are estimated for $p \in \{0, 1, \ldots, 10\}$ and the model with the smallest BIC is chosen. The strategies for model selection of $ARFIMA(p,d,q)$ models involve estimation of $(P+1)(Q+1)$ models where P and Q are the maximum orders of the short memory parameters being considered. They were generally fixed at 8 requiring estimation of 81 models and the model with the smallest BIC is chosen. Robust standard errors are in parentheses. $ln(L)$ represents the maximized log-likelihood and BIC represents the Bayesian information criterion.

complements the linear $ARFIMA$ model estimation in Table 3.2. We report estimates of the long memory parameter from two semi parametric procedures. First, the Local Whittle (LW) estimator, which is obtained by minimizing the objective function,

$$R^{LW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d} I_y(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j),$$

with respect to d, where $\omega_j = (2\pi j)/T$ for $j = 1, 2, ..., T$ and $I_y(\omega_j)$ is the periodogram defined as,

$$I_y(\omega_j) = \frac{1}{2\pi T}\left|\sum_{j=1}^{T} y_t e^{i\omega_j t}\right|^2.$$

The estimator depends on the choice of bandwidth, m, which is generally chosen as $m = \lfloor T^\delta \rfloor$ where $0 < \delta < 4/5$; and where $\lfloor\cdot\rfloor$ denotes the integer part. Several important extensions of the LW estimator have been introduced in the literature. In particular, Shimotsu and Phillips (2005) have proposed the Exact Local Whittle

(ELW) approach using a "corrected" discrete Fourier transform of the series, where the objective function now becomes,

$$R^{ELW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} I_{\nabla^d y}(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j),$$

where $\nabla^d = (1-L)^d$. Given the distinct possibility of non stationary long memory RV series we also use the method of Abadir *et al.* (2007), who have introduced the Fully Extended Local Whittle ($FELW$) where $d \in (p-1/2, p+1/2]$, for $p = 0, 1, 2, ...,$ which has the particular attraction of covering the region of nonstationarity for long memory processes. Then,

$$I^{FELW}(\omega_j) = \left|1 - e^{i\omega_j}\right|^{-2p} I_{\nabla^p y}(\omega_j),$$

where the $FELW$ is obtained by minimizing,

$$R^{FELW}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} j^{2d} I^{FELW}(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(j).$$

The LW is known to be a consistent estimator of d in the stationary region of $-1/2 < d < 1/2$ with $m^{1/2}\left(\widehat{d_{LW}} - d_0\right) \to N\left\{0, (1/4)\right\}$. While the ELW and $FELW$ estimators are known to be consistent for all values of d.

A particularly important issue concerns the choice of bandwidth, denoted by m, which is generally chosen in the range of $T^{1/2} \leq m \leq T^{4/5}$. The LW and $FELW$ statistics are also reported in Table 3.2 and similarly to the time domain methods they find very significant long memory features of the RV series. However, both the LW and $FELW$ statistics are very dependent on the choice of bandwidth, m. For this reason the LW and $FELW$ estimators are also reported for a selection of bandwidth choices; including $m = T^{0.5}$, which tends to be the conventional choice, and also $m = T^{0.3}$ and $m = T^{0.7}$. The latter gives considerably more weight to the short frequency components that are apparently of importance as evidenced by the need for relatively large number of short memory parameters selected in the $ARFIMA$ estimation and also the HAR model considered later.

Overall, there is clear evidence of long memory characteristics from the $ARFIMA$ estimation and also the complementary LW and $FELW$ semi parametric results. The overall results suggest that the estimated long memory is either inside, or very close to the region of nonstationarity.

It is possible that the very significant estimates of the long memory parameter are due to non-linear effects, or due to structural breaks. In particular, Granger and Hyung (2004) show that occasional break points processes are hard to distinguish from a pure fractional, $I(d)$ model. Alternative non-linear explanations have centred on the possibility of regime switches giving rise to the appearance of long memory in a time series; see Granger and Ding (1996), Granger and Teräsvirta (1999) and particularly Diebold and Inoue (2001) who showed that a Markov Switching regime change model that can generate a long memory time series. Davidson and Sibbertsen (2005) discuss other regime switching and non-linear models which can generate long memory. For these reasons we also used the tests of Sibbertsen (2004) and Wenger *et al.* (2018), who have provided a $CUSUM$ test to test for structural breaks in the intercept of long memory process. This change in mean test is used to check the robustness of the long memory hypothesis; and on applying this test to the fractionally filtered series, it is denoted as $CUSUM- \nabla^d$. The test statistic is defined as

$$Q_T = \sup_{r\in(0,1)} \left| \left(\widehat{\sigma^2} T\right)^{-1/2} \sum_{t=1}^{[rT]} \widehat{u}_t^* \right|,$$

where $(1-L)^{\widehat{d}} y_t = y_t^*$, and $\widehat{d}$ is the LW or $FELW$. Furthermore $\widehat{u}_t^* = y_t^* - \overline{y^*}$, where $\overline{y^*} = \frac{1}{T}\sum_{t=1}^{T} y_t^*$ and $\widehat{\sigma^2} = \frac{1}{T}\sum_{j=1}^{m} \widehat{u}_t^{*2}$. Wenger *et al.* (2018) show that the limiting distribution of Q_T is pivotal with respect to $\widehat{d}$ and that the test statistic follows the conventional distribution as defined by Ploberger and Krämer (1992). The critical values for Q_T at the $0.01, 0.05$ and 0.10 significance levels are 1.63, 1.36 and 1.22 respectively. The $CUSUM$ statistics results indicate that the only RV series with some evidence for structural change in the mean are the Australian dollar and the $S\&P500$.

TABLE 3.2: Estimates of Long Memory Parameter d using *LW* and *FELW*

		AUD	CAD	EUR	GBP	JPY	S&P500
		LW ($m = \lfloor T^b \rfloor$)					
$b = 0.3$	d	0.316	0.606	0.385	0.381	0.541	0.312
		(0.146)	(0.146)	(0.146)	(0.146)	(0.146)	(0.143)
0.5	d	0.495	0.731	0.756	0.779	0.501	0.473
		(0.270)	(0.270)	(0.270)	(0.270)	(0.270)	(0.268)
0.7	d	0.678	0.598	0.558	0.643	0.380	0.734
		(0.325)	(0.325)	(0.325)	(0.325)	(0.325)	(0.323)
		FELW ($m = \lfloor T^b \rfloor$)					
$b = 0.3$	d	0.274	0.495	0.294	0.344	0.390	0.245
		(0.146)	(0.146)	(0.146)	(0.146)	(0.146)	(0.143)
0.5	d	0.501	0.708	0.743	0.794	0.510	0.474
		(0.270)	(0.270)	(0.270)	(0.270)	(0.270)	(0.268)
0.7	d	0.606	0.564	0.518	0.556	0.336	0.681
		(0.325)	(0.325)	(0.325)	(0.325)	(0.325)	(0.323)
$b = 0.5$	*CUSUM*-$\bigtriangledown^d$	2.020	0.671	0.546	0.674	1.067	1.912

Note: The *LW* and *FELW* estimators are estimated with bandwidths (m)= $\lfloor T^b \rfloor$ with $b \in \{0.3, 0.5, 0.7\}$. Robust standard errors are reported in parentheses. *CUSUM*-$\bigtriangledown^d$ statistic is the usual *CUSUM* statistic applied to the d (estimated with *FELW*, $m = T^{0.5}$) fractionally filtered series.

3.6 HAR Models

The HAR models require defining h period averages of the observed RV series

$$\overline{RV}_{t,t+h} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+i},$$

where $h = 1, 5,$ and 22 for the one day, one week, and one month cumulative volatilities. This three parameter HAR model is motivated by the additive partial cascade of volatilities model. On further defining $\overline{RV}_t^w = \frac{1}{5}\sum_{j=0}^{4} RV_{t-j}$ as the weekly average, and $\overline{RV}_t^m = \frac{1}{22}\sum_{j=0}^{21} RV_{t-j}$ as the monthly average; then the HAR model reduces to

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d RV_t + \left(\frac{\phi_w}{5}\right)\sum_{i=0}^{4} RV_{t-i} + \left(\frac{\phi_m}{22}\right)\sum_{i=0}^{21} RV_{t-i} + \varepsilon_{t+h} \tag{3.3}$$

which is a restricted parameter version of the general $AR(22)$ model and is represented as,

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d \; RV_t + \phi_w \overline{RV}_t^w + \phi_m \overline{RV}_t^m + \varepsilon_{t+h}. \tag{3.4}$$

The original model was proposed by Corsi (2009) to explain the persistence of RV series from the heterogeneity of an agent's behavior over distinct time horizons. The HAR model is generally described as an additive volatility cascade, from high frequencies to low frequencies; with each additive cascade having close to an $AR(1)$ structure. The notion of multiple components in the volatility process is justified in terms of differences of agents risk profiles, institutional structures, temporal horizons, etc.

Table 3.3: Estimation of the basic *HAR* Model

	$\overline{RV}_{h,t+h} = \phi_0 + \phi_d RV_t^{(d)} + \phi_w \overline{RV}_t^{(w)} + \phi_m \overline{RV}_t^{(m)} + \varepsilon_{t+h}$					
	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_d	0.415	0.270	0.272	0.077	0.223	0.222
	(0.070)	(0.083)	(0.052)	(0.054)	(0.081)	(0.122)
ϕ_w	0.119	0.275	0.244	0.145	0.197	0.330
	(0.087)	(0.096)	(0.069)	(0.069)	(0.065)	(0.144)
ϕ_m	0.343	0.370	0.401	0.542	0.364	0.337
	(0.069)	(0.070)	(0.057)	(0.060)	(0.061)	(0.106)
$ln(L)$	–3001.364	28.371	48.230	–1266.077	–2191.938	–6376.967
BIC	6043.708	–15.762	–55.479	2573.134	4424.857	12795.614

Note: *OLS* estimates of the basic *HAR* model are reported with robust standard errors in parentheses. $ln(L)$ is the maximized log-likelihood.

Table 3.4: Estimation of the *EHAR* Model

	$\overline{RV}_{h,t+h} = \phi_0 + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \phi_w \overline{RV}_t^{(w)} + \phi_m \overline{RV}_t^{(m)} + \varepsilon_{t+h}$					
	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_d^+	0.601	0.355	0.285	–0.079	0.179	–0.009
	(0.150)	(0.107)	(0.086)	(0.019)	(0.166)	(0.194)
ϕ_d^-	0.157	0.185	0.260	0.550	0.252	0.425
	(0.137)	(0.102)	(0.069)	(0.189)	(0.163)	(0.177)
ϕ_w	0.140	0.278	0.242	0.106	0.202	0.352
	(0.086)	(0.084)	(0.070)	(0.059)	(0.064)	(0.145)
ϕ_m	0.354	0.367	0.401	0.462	0.365	0.333
	(0.068)	(0.070)	(0.057)	(0.072)	(0.060)	(0.105)
$ln(L)$	–2979.832	32.935	48.333	–1133.620	–2191.037	–6344.528
BIC	6008.841	–16.694	–47.489	2316.417	4431.251	12739.072

Note: As in Table 3.3 with ordinary least square estimates of the *HAR* model with in addition positive and negative semi variances reported, or *EHAR*.

Estimates of the HAR model are reported in Table 3.3 for the six RV series. The OLS estimates of the parameters are largely consistent with those of previous studies with the estimated daily HAR parameter, ϕ_d being statistically significant and in the range of 0.22 to 0.42 for five of the RV series. The value of the estimated ϕ_w varies substantially across series and is generally statistically significant. The estimated ϕ_m parameter is in the range of 0.33 to 0.54 and is very significant for all the six RV series. On comparing the estimated $ARFIMA$ models in Table 3.1 and the estimated HAR models in Table 3.3, it is clear that the $ARFIMA$ models dominate the HAR models in terms of BIC model selection[1].

The basic HAR model has been extended by Patton and Sheppard (2015) to include separate effects of volatility due to positive and negative returns and to include good and bad volatility through the signed jump variation. We use the usual terminology of denoting such models as Extended HAR, or simply $EHAR$, Table 3.4.

TABLE 3.5: Estimation of the *EHAR* Model (cont'd)

	$\overline{RV}_{h,t+h} = \phi_0 + \phi_J^+ \Delta J_t^{2+} + \phi_J^- \Delta J_t^{2-} + \phi_C BV_t + \phi_w \overline{RV}_t^{(w)} + \phi_m \overline{RV}_t^{(m)} + \varepsilon_{t+h}$					
	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_J^+	0.362	0.017	–0.060	–0.085	–0.325	0.277
	(0.308)	(0.163)	(0.089)	(0.040)	(0.053)	(0.241)
ϕ_J^-	0.029	0.128	0.031	–0.135	0.312	–0.749
	(0.176)	(0.075)	(0.090)	(0.137)	(0.184)	(0.295)
ϕ_C	0.437	0.443	0.420	0.222	0.590	0.149
	(0.107)	(0.071)	(0.096)	(0.150)	(0.074)	(0.147)
ϕ_w	0.104	0.188	0.165	0.111	0.065	0.327
	(0.098)	(0.087)	(0.075)	(0.066)	(0.047)	(0.140)
ϕ_m	0.355	0.349	0.379	0.481	0.281	0.309
	(0.069)	(0.071)	(0.061)	(0.083)	(0.055)	(0.094)
$ln(L)$	–2939.628	100.703	97.773	–1159.155	–2059.240	–6256.314
BIC	5936.630	–144.033	–138.174	2375.682	4175.853	12570.980

Note: As in Table 3.3 with ordinary least square parameter estimates and robust standard errors of the *EHAR* model with positive and negative signed variation and BV reported.

In terms of BIC, the $ARFIMA(p,d,0)$ model is preferred to the HAR and extended HAR models with semi variances, for all assets except Canada. While the $ARFIMA(p,d,0)$ is preferred to $EHAR$ with jumps for all but the Canadian dollar and the Euro. While the $ARFIMA(p,d,q)$ model is better than HAR and $EHAR$ with semi variances for all assets; and the $ARFIMA(p,d,q)$ is preferred to $EHAR$ with jumps for all but the Canadian dollar. In summary, when the HAR, or extended HAR are combined with long memory the estimate of d is significant. However, the

[1]McAleer and Medeiros (2008) have considered an alternative model formulation which combines smooth transition regimes and long range dependence.

$ARFIMA(p, q, 0)$ and $ARFIMA(p, d, q)$ models were preferred over HAR or extended HAR in the majority of cases.

Table 3.5 reports estimates of another version of the $EHAR$ model which supplement the terms in the basic HAR model to include signed semivariances which distinguish between positive and negative returns (in Equation (3.5)) and separate positive and negative signed jumps as in Equation (3.1) with BV (in Equation (3.6)). These models are

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_d^+ \ RS_t^+ + \phi_d^- \ RS_t^- + \phi_w \overline{RV}_t^w + \phi_m \overline{RV}_t^m + \varepsilon_{t+h} \tag{3.5}$$

and

$$\overline{RV}_{t,t+h} = \phi_0 + \phi_J^+ \Delta J_t^{2+} + \phi_J^- \Delta J_t^{2-} + \phi_C BV_t + \phi_w \overline{RV}_t^w + \phi_m \overline{RV}_t^m + \varepsilon_{t+h}. \tag{3.6}$$

The last two models were all introduced in Patton and Sheppard (2015). Similar jump extended HAR models were studied by Andersen *et al.* (2007) and Busch *et al.* (2011). Exactly the same conclusions emerge from a comparison of the estimated $ARFIMA$ models in Table 3.1 with the extended HAR model in Table 3.5, where the daily RV component is omitted and replaced with RS^+ and RS^- respectively. The ϕ_C parameter is associated with the "continuous" BV, which is intended to make the continuous part of RV robust to the presence of the jumps. If there are no jumps, then daily RV should be asymptotically identical to BV. Good jumps lead to lower volatility, and bad jumps lead to higher volatility in longer horizons. There is evidence in the second panel of Table 3.5 that the presence of the signed jump variables are effective in explaining the RV for Japan, but not for the other RV exchange rate series. However, the parameter associated with the negative jump variable is highly and negatively significant for the $S\&P500$ RV series. Similar results have been found by Busch *et al.* (2011).

However, the parameter associated with Bi-Power Variation BV_t in Table 3.5 is highly significant; while the estimated ϕ_m parameter is statistically significant and between 0.28 and 0.48 across the various assets RV series. The importance of the essentially $AR(22)$ term seems to indicate the need for higher order dynamics or for long memory. This possibility is pursued in the next section.

3.7 Distinguishing HAR from long memory

So far we have presented favourable evidence for the presence of long memory and also for the validity of the HAR model. While BIC generally favours the long memory models over HAR it is also worth further investigation to try to distinguish between these two theories. One issue with empirical work in this area is to accurately distinguish between long memory and very persistent short memory autoregressive behaviour. This problem becomes particularly apparent in the high correlation between the estimated long memory parameter and the estimated short memory $ARMA$ parameters and resulting instability of these parameter estimates in the presence of higher order parametrizations. The same problem is apparent in the frequency domain LW and $FELW$ where the choice of bandwidth is so critical and the noted poor performance of these semi parametric estimators in the presence of very persistent autocorrelation; e.g. see Baillie and Kapetanios (2008) and Nielsen and Frederiksen (2005).

In this section we tackle these problems in several different directions. First, we estimate HAR models from simulated long memory processes and tabulate the properties of the resulting simulated HAR parameter estimates. Second, we estimate both unrestricted $ARFIMA(22, d, 0)$ models and restricted $ARFIMA$ models, (denoted as $RARFIMA$), where the parameter restrictions are implied by the HAR model. Third, we estimate by MLE a similar theoretical model which estimates HAR models with long memory disturbances. All of these methods provide different pieces of evidence on the issue of distinguishing one model, or property, from another.

3.7.1 Simulating Estimated HAR models from a long memory process

The first approach is to generate realizations from $ARFIMA(0, d, 0)$ model with different long memory parameters of $d \in \{0.25, 0.30, 0.35, 0.40, 0.45\}$. Each generated series has $T = 10,000$ observations and we perform $5,000$ replications for each design, to estimate the basic three parameter HAR models. It should be noted that simulation of a data series of this magnitude can be greatly facilitated by using the fast fractional differencing algorithm of Jensen and Nielsen (2014). For a series of $T = 10,000$ observations, their algorithm which is exact, and not an approximation, is more than 15 times faster than the standard linear convolution implantation in $MATLAB$; see

Table 1 of Jensen and Nielsen (2014).

In many respects the simulation results in Table 3.6 replicate many of the features of HAR estimation in this paper and other literature. The mean of the simulated estimated ϕ_d parameter is 0.20 for a data generating process of $ARFIMA(0, 0.25, 0)$ and increases monotonically as d increases to 0.41 for when the simulated series is from an $ARFIMA(0, 0.45, 0)$ process. The interval 0.37 to 0.44 provides a 95% coverage of the monthly HAR parameter ϕ_d from an $ARFIMA(0, 0.45, 0)$ design and so appears relatively precise. Similar degrees of precision are found for the other simulated parameters. However, ϕ_w and ϕ_m have much less variation with the value of d and lie in the range of 0.23 to 0.29 for all cases.

The above results can be compared with those in Table 3.4, which have many similar features; although GBP has considerably lower ϕ_d than predicted and rather higher ϕ_m than predicted.

3.7.2 Restricted ARFIMA Models

The second line of investigation focuses on using MLE to estimate both unrestricted $ARFIMA(22, d, 0)$ models and restricted version of the model, an $RARFIMA(22, d, 0)$. The parameter restrictions on this latter model are those implied in Equation (3.3). Hence the $RARFIMA(22, d, 0)$ model is

$$(1 - L)^d \lambda(L) RV_t = \varepsilon_t, \tag{3.7}$$

where $\lambda(L) = 1 - \lambda_1 L - \lambda_2 L^2 - \lambda_2 L^3 - \lambda_2 L^4 - \lambda_2 L^5 - \lambda_3 L^6 - \lambda_3 L^7 \cdots - \lambda_3 L^{22}$ with all the roots of $\lambda(L)$ outside the unit circle and ε_t denoting white noise. This model is identical to $\phi(L)(1 - L)^d RV_t = \varepsilon_t$ with 19 restrictions, $\phi_2 = \phi_3 = \phi_4 = \phi_5 \equiv \lambda_2$ and $\phi_6 = \phi_7 = \cdots = \phi_{22} \equiv \lambda_3$. The MLE of the $RARFIMA(22, d, 0)$ model parameters are to be found in Table 3.7. Results for the unrestricted $ARFIMA(22, d, 0)$ models are not presented and generally possess correlations between the estimated long memory parameter and the twenty two unrestricted autoregressive parameters. Likelihood Ratio (LR) tests of the 19 parameter restrictions which reduce the unrestricted $ARFIMA(22, d, 0)$ model to the $RARFIMA(22, d, 0)$ are presented in Table 3.7. The LR tests reject the restrictions that are consistent with a HAR model for all the RV series. From Table 3.7 it can be seen that the MLE of the long memory parameter d

is around 0.30 for four of the RV series and is not significantly different from zero for the Euro or the $S\&P500$ RV series. In general these results suggest that the HAR model provides a useful representation of some of the low order dynamics of RV, but that long memory also plays an important role to describe higher order dynamics.

TABLE 3.6: Simulated HAR Estimations from Fractional White Noise

	$d = 0.25$			$d = 0.30$		
	ϕ_d	ϕ_w	ϕ_m	ϕ_d	ϕ_w	ϕ_m
Mean($\hat{\phi}$)	0.204	0.227	0.228	0.252	0.252	0.235
SD($\hat{\phi}$)	0.017	0.033	0.048	0.017	0.032	0.044
Mean(se($\hat{\phi}$))	0.016	0.031	0.040	0.016	0.030	0.035
	$d = 0.35$			$d = 0.40$		
	ϕ_d	ϕ_w	ϕ_m	ϕ_d	ϕ_w	ϕ_m
Mean($\hat{\phi}$)	0.302	0.270	0.232	0.354	0.281	0.222
SD($\hat{\phi}$)	0.017	0.030	0.040	0.018	0.029	0.036
Mean(se($\hat{\phi}$))	0.017	0.029	0.031	0.016	0.027	0.027
	$d = 0.45$					
	ϕ_d	ϕ_w	ϕ_m			
Mean($\hat{\phi}$)	0.407	0.285	0.206			
SD($\hat{\phi}$)	0.017	0.028	0.031			
Mean(se($\hat{\phi}$))	0.016	0.026	0.024			

Note: For each panel, Mean($\hat{\phi}$) is the average value of each estimated HAR parameters across 5,000 iterations. Similarly, SD($\hat{\phi}$) is the standard deviation of those estimates and Mean(se($\hat{\phi}$)) refers to the average standard error of those estimates.

A related method to the above is to specify the long memory process as a disturbance around the HAR specification and to estimate the model

$$(1-L)^d \left(y_t - x_t'\beta\right) = \varepsilon_t, \qquad t = 1, \ldots, T, \tag{3.8}$$

where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E(\epsilon_t \epsilon_s) = 0$, $s \neq t$. While x_t is a k dimensional vector of explanatory HAR type variables at time t, and β is the corresponding vector of parameters.

The implementation of MLE to the above model follows as in Section 3.5. If all the variables are $I(d)$ with $-0.5 < d < 0.5$ then conventional asymptotic are valid and the MLE should be $T^{1/2}$ consistent. However, when HAR variables are included in the

Table 3.7: Estimation of the $RARFIMA(22, d, 0)$ model

	AUD	CAD	EUR	GBP	JPY	S&P500
λ_1	0.317	0.014	–0.163	–0.006	0.054	0.387
	(0.121)	(0.116)	(0.037)	(0.234)	(0.102)	(1.171)
λ_2	0.045	0.044	–0.060	0.044	–0.008	0.062
	(0.025)	(0.043)	(0.022)	(0.076)	(0.025)	(0.120)
λ_3	0.016	0.031	0.004	0.022	0.015	0.015
	(0.008)	(0.006)	(0.012)	(0.011)	(0.013)	(0.048)
σ^2	0.305	0.058	0.057	0.089	0.187	1.142
	(0.049)	(0.006)	(0.006)	(0.021)	(0.033)	(0.199)
d	0.305	0.310	0.518	0.364	0.362	0.118
	(0.119)	(0.120)	(0.325)	(0.264)	(0.085)	(1.167)
$ln(L)$	–2993.725	10.280	50.416	–756.124	–2105.726	–6196.369
BIC	6028.431	20.421	–51.655	1553.229	4252.434	12434.419
Wald	48.343	35.799	75.575	47.664	25.857	25.814

Note: The $RARFIMA(22, d, 0)$ model is $\lambda(L)(1-L)^d(RV_t-\mu)=\varepsilon_t$ where $\lambda(L)=1-\lambda_1L-\lambda_2L^2-\lambda_2L^3-\lambda_2L^4-\lambda_2L^5-\lambda_3L^6-\lambda_3L^7-\lambda_3L^8-\cdots-\lambda_3L^{22}$. This model is identical to $\phi(L)(1-L)^d(RV_t-\mu)=\varepsilon_t$ with 19 restrictions, $\phi_2=\phi_3=\phi_4=\phi_5\equiv\lambda_2$ and $\phi_6=\phi_7=\cdots=\phi_{22}\equiv\lambda_3$. In the last column, the Wald statistic is computed from the unrestricted $ARFIMA(22, d, 0)$ model with these 19 HAR restrictions as the null hypothesis. The 1% and 5% critical values for χ^2_{19} distribution are 43.82 and 35.58, respectively. Robust standard errors are in parentheses.

Table 3.8: Estimation of the *HAR* Model with Long Memory Error Process

	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_d	0.229	0.144	0.097	0.034	0.065	–0.023
	(0.042)	(0.168)	(0.060)	(0.012)	(0.083)	(0.076)
ϕ_w	0.035	0.170	0.077	0.050	0.071	–0.033
	(0.133)	(0.194)	(0.105)	(0.037)	(0.081)	(0.202)
ϕ_m	0.442	0.518	0.517	0.244	0.238	0.364
	(0.180)	(0.195)	(0.103)	(0.101)	(0.171)	(0.384)
d	0.298	0.146	0.239	0.365	0.295	0.478
	(0.110)	(0.173)	(0.074)	(0.060)	(0.096)	(0.152)
$ln(L)$	–2937.37	35.832	79.35	–732.11	–2091.54	–6204.68
BIC	5923.92	–22.486	–109.52	1513.39	4232.25	12459.37

Note: Approximate *MLE*s of the *HAR* model with *ARFIMA* errors reported. *QMLE* standard errors are in parentheses.

regression, there is the possibility of some variables having $d > 0.5$ and hence being non stationary long memory processes; and also the possibility of forms of non standard fractional cointegration occurring.

The long memory property of RV is a feature shared by many other volatility series, which gives rise to the possibility of fractional cointegration between volatility series. This has been considered by Christensen and Nielsen (2006) and Bollerslev *et*

TABLE 3.9: Estimation of the *ARFIMA-EHAR* Model

	ARFIMA-EHAR model with positive and negative semivariances					
	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_d^+	0.374	0.214	0.061	0.034	0.047	–0.211
	(0.134)	(0.222)	(0.099)	(0.031)	(0.079)	(0.147)
ϕ_d^-	0.074	0.111	0.120	0.033	0.077	0.137
	(0.133)	(0.190)	(0.069)	(0.086)	(0.122)	(0.128)
ϕ_w	0.058	0.192	0.075	0.050	0.074	–0.023
	(0.142)	(0.202)	(0.107)	(0.037)	(0.079)	(0.521)
ϕ_m	0.448	0.495	0.516	0.244	0.240	0.370
	(0.175)	(0.221)	(0.105)	(0.103)	(0.176)	(0.376)
d	0.272	0.124	0.244	0.366	0.294	0.487
	(0.114)	(0.192)	(0.076)	(0.058)	(0.097)	(0.254)
$ln(L)$	–2926.69	37.38	80.00	–732.11	–2091.36	–6171.65
BIC	5910.75	–17.38	–102.62	1521.58	4240.09	12401.65

Note: Similar to Table 3.8 with approximate *MLE*s of the *ARFIMA-EHAR* model with positive and negative semivariances reported.

al. (2013), who include the possibility that volatility predicts returns. In fact, several articles consider this in the context of long memory models; see Christensen and Nielsen (2007) and Christensen *et al.* (2010) who use a $FIEGARCH-M$ model which builds on the $FIGARCH$ model of Baillie *et al.* (1996) and $FIEGARCH$ model of Bollerslev and Mikkelsen (1996). The latter paper deals with differences from positive and negative returns and is particularly relevant in the context of realized semi variances and jump variation in RV. One attraction of high frequency data is that it is model free and does not require the formulation of Stochastic Volatility or $GARCH$ type models. The interest in the HAR approach is that it provides possible alternatives to long memory and fractional cointegration analysis.

The potential borderline non stationarity of long memory of the RV series creates particularly challenging problems and is not pursued in this study. Hence while we believe the estimates are likely consistent, we note that there is some uncertainty regarding the validity of the conventional standard error estimates being reported. The resulting estimated models are reported in Tables 3.9 and 3.10 with many of the ϕ_d estimates being particularly significant across assets RV series; while the relative importance of the other volatility parameters ϕ_c and ϕ_w being less statistically important. Of particular interest is the magnitude and significance of the long memory parameter d, which is highly statistically significant across all RV series except for the Canadian dollar.

TABLE 3.10: Estimation of the *ARFIMA-EHAR* model (cont'd)

	ARFIMA-EHAR model with positive and negative signed variations and BV					
	AUD	CAD	EUR	GBP	JPY	S&P500
ϕ_J^+	0.201	0.067	–0.139	0.030	–0.262	0.104
	(0.317)	(0.218)	(0.095)	(0.025)	(0.077)	(0.197)
ϕ_J^-	0.073	0.107	0.054	–0.087	0.270	–0.469
	(0.183)	(0.088)	(0.078)	(0.192)	(0.132)	(0.304)
ϕ_C	0.305	0.473	0.278	0.048	0.416	–0.053
	(0.079)	(0.113)	(0.167)	(0.020)	(0.121)	(0.090)
ϕ_w	0.037	0.218	0.087	0.040	0.038	0.030
	(0.161)	(0.098)	(0.111)	(0.034)	(0.059)	(0.448)
ϕ_m	0.440	0.300	0.481	0.246	0.283	0.424
	(0.161)	(0.132)	(0.112)	(0.102)	(0.082)	(0.255)
d	0.236	–0.050	0.155	0.361	0.159	0.413
	(0.114)	(0.092)	(0.137)	(0.062)	(0.090)	(0.215)
$ln(L)$	–2897.95	102.06	109.16	–724.77	–2039.08	–6142.23
BIC	5861.47	–138.55	–152.75	1515.12	4143.73	12351.14

Note: Similar to Table 3.8 with approximate *MLE*s of the *ARFIMA-EHAR* model with positive and negative signed variations and BV reported.

One relevant comparison is the basic HAR estimation in Table 3.4 with the estimated above model from Equation 3.8 in Table 3.8. The estimated long memory parameter is significant for five of the six RV series and the BIC prefer the model with long memory to the basic HAR formulation. Interestingly there appears to be less role for the long memory parameter when estimating the extended HAR model with ϕ_c and signed jump variables. Also, the impact of the negative signed jump variable for the $S\&P500$ series still has a negatively signed parameter estimate and is now not significant. This provides an interesting comparison with the results in Table 3.5 and suggests that the presence of a jump variable maybe picking up discontinuities which are otherwise giving rise to the presence of long memory.

3.8 Time varying parameter extended HAR models

While long memory appears to be an important modelling feature to include with the extended HAR formulation, another possibility worth considering is that the weights attached to the partial cascade volatilities may not be constant over time. If this were the case, the non linearity may well capture some of the long memory aspects of the RV series. If we view the partial cascade volatilities reflecting agent's risk preferences, access to information, geographical location, etc then there are many explanations why the coefficients may have time variation. On modifying Equation (3.4), we can write the $TVP-HAR$ model as

$$\overline{RV}_{t,t+h} = \phi_{0,t} + \phi_{d,t}\ RV_t + \phi_{w,t}\overline{RV}_t^w + \phi_{m,t}\overline{RV}_t^m + \varepsilon_{t+h}$$

where the $\phi_{j,t}$ coefficients are now time varying and are partial volatility parameters that depend on time varying risk premium. This model is implemented as a kernel weighted regression which is facilitated by an extension of the random coefficient approach of Giraitis *et al.* (2014).

Details of the means and standard deviations of the estimated parameters in the $TVP-HAR$ model for all RV assets are presented in Table 3.11. Similar results for the estimated $TVP-EHAR$ models are available in Table 3.12. Showing the first two moments of these parameter estimates as they change over time is only part of the story and some idea about their variability can be seen in Figures 3.3 and 3.4 which plot the time variation of the parameter estimates across the sample. We only present figures for the time variation in the parameters for the Australian dollar and the Euro vis a vis the US dollar to conserve space. Full details of the figures for the remaining assets are available from the author on request. A potentially interesting topic for future research would be to see if the variation in the parameter estimates are related to particular economic episodes.

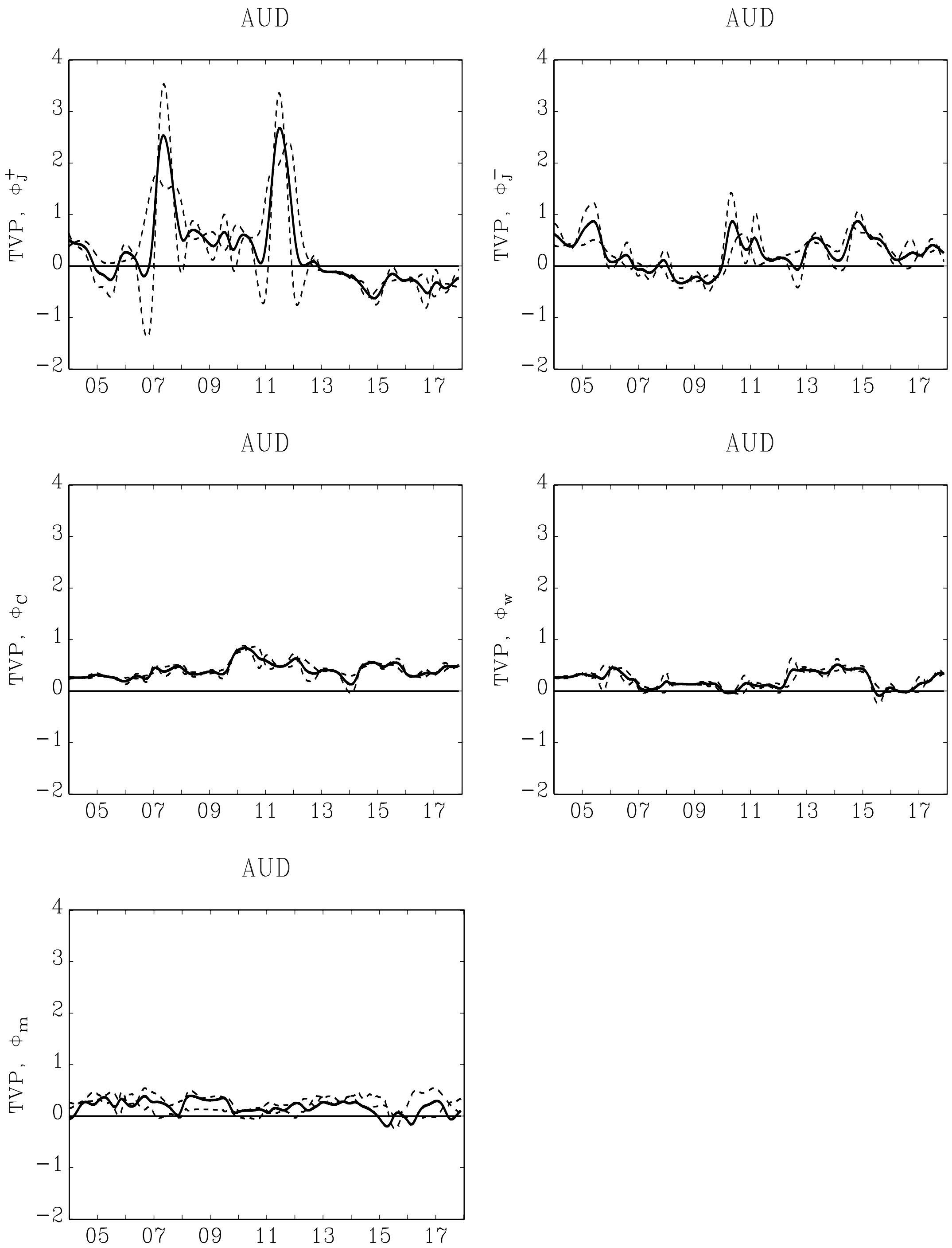
However, in terms of model comparisons using BIC, the $RARFIMA$ model in Equa-

Table 3.11: Parameter estimation of the *TVP-HAR* Model

	AUD	CAD	EUR	GBP	JPY	S&P500
$\phi_{d,t}$	0.287	0.166	0.250	0.189	0.246	0.293
	(0.196)	(0.179)	(0.186)	(0.138)	(0.165)	(0.189)
$\phi_{w,t}$	0.260	0.316	0.180	0.281	0.208	0.308
	(0.206)	(0.155)	(0.142)	(0.182)	(0.191)	(0.186)
$\phi_{m,t}$	0.178	0.256	0.272	0.226	0.166	0.120
	(0.127)	(0.125)	(0.168)	(0.161)	(0.149)	(0.133)
BIC	7791.801	1679.234	1498.431	3241.878	5876.518	14621.189

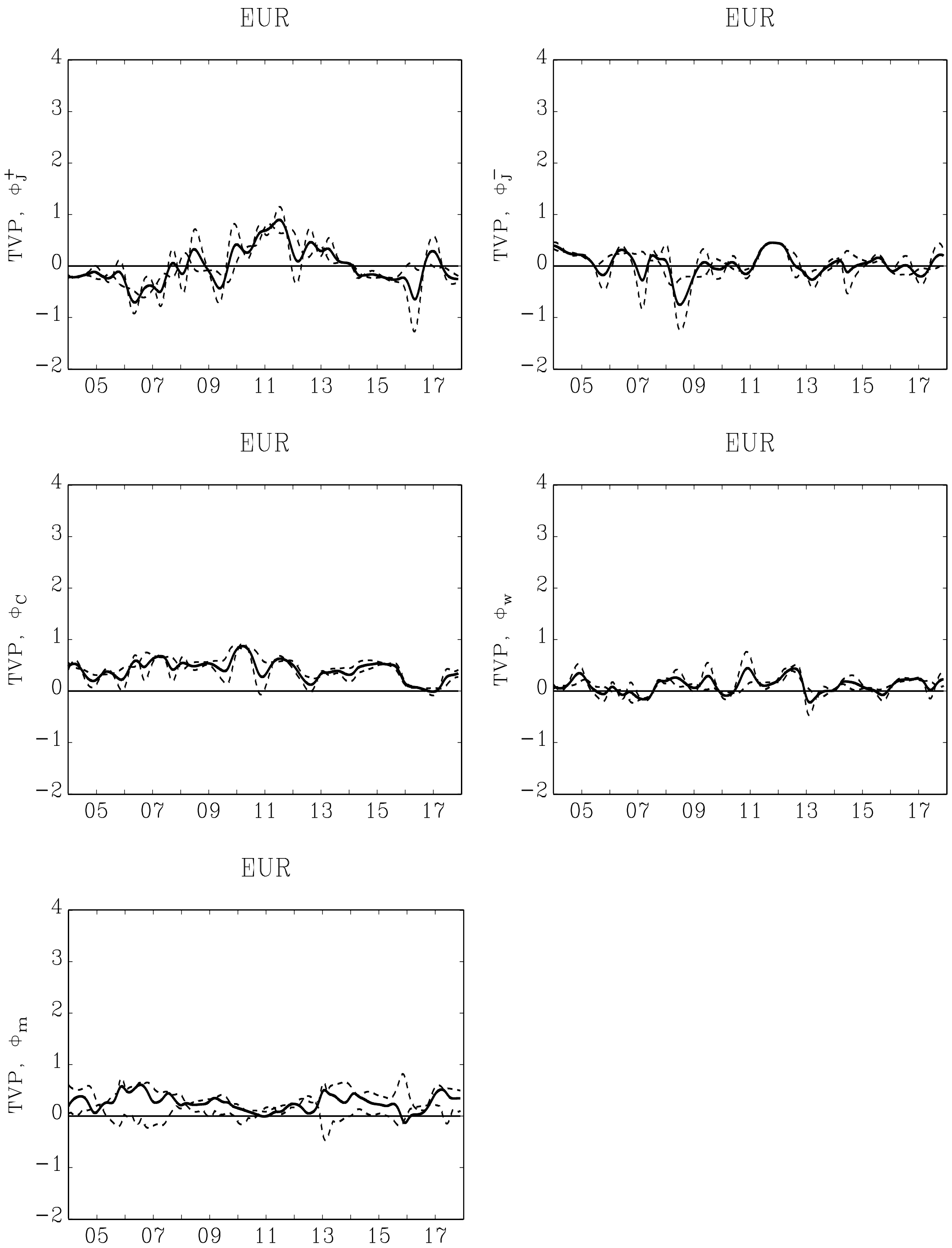
Note: The mean values of the coefficients of the *TVP-HAR* are reported with standard deviation in parentheses. The Gaussian kernel with a bandwidth of $T^{0.5}$ is used for estimation.

tion (3.7), which includes both HAR parameters and long memory; and the related



Note: The figure plots the time variation of the parameter estimates of the $TVP - EHAR$ model across the sample for the Australian dollar. Full details of the figures for the remaining assets are available from the author on request.

FIGURE 3.3: Estimation results from the *TVP-EHAR* model - Australian Dollar



Note: The figure plots the time variation of the parameter estimates of the $TVP - EHAR$ model across the sample for the Euro. Full details of the figures for the remaining assets are available from the author on request.

FIGURE 3.4: Estimation results from the *TVP-EHAR* model - Euro

model in Equation (3.8) outperforms the $TVP-HAR$ alternatives in terms of BIC. Hence the $RARFIMA$ based models with constant long memory parameter and constant HAR parameters appear preferable based on model selection criteria.

Table 3.12: Parameter estimation of the *TVP-EHAR* Model

	AUD	CAD	EUR	GBP	JPY	S&P500
	TVP-EHAR model with positive and negative semivariances					
$\phi^+_{d,t}$	0.555	0.307	0.290	0.226	0.128	0.024
	(0.507)	(0.328)	(0.309)	(0.222)	(0.375)	(0.238)
$\phi^-_{d,t}$	–0.002	0.033	0.221	0.168	0.358	0.553
	(0.273)	(0.185)	(0.238)	(0.232)	(0.377)	(0.261)
$\phi_{w,t}$	0.269	0.311	0.178	0.278	0.223	0.322
	(0.190)	(0.141)	(0.136)	(0.182)	(0.190)	(0.188)
$\phi_{m,t}$	0.187	0.257	0.270	0.225	0.163	0.129
	(0.128)	(0.123)	(0.165)	(0.157)	(0.145)	(0.133)
BIC	8140.775	2078.796	1936.916	3693.370	6276.771	15044.607
	TVP-EHAR model with positive and negative signed variations and BV					
$\phi^+_{J,t}$	0.281	0.046	–0.004	0.048	–0.201	–0.062
	(0.741)	(0.339)	(0.355)	(0.260)	(0.463)	(0.250)
$\phi^-_{J,t}$	0.244	0.200	0.020	–0.035	0.020	–0.445
	(0.310)	(0.329)	(0.225)	(0.424)	(0.617)	(0.412)
$\phi_{C,t}$	0.397	0.333	0.398	0.311	0.445	0.331
	(0.149)	(0.189)	(0.185)	(0.141)	(0.324)	(0.227)
$\phi_{w,t}$	0.213	0.231	0.117	0.230	0.158	0.288
	(0.160)	(0.138)	(0.147)	(0.163)	(0.178)	(0.195)
$\phi_{m,t}$	0.176	0.247	0.257	0.212	0.152	0.134
	(0.131)	(0.129)	(0.155)	(0.147)	(0.154)	(0.127)
BIC	8496.352	2409.951	2345.089	4046.409	6635.650	15317.870

Note: The mean values of the coefficients of the *TVP-EHAR* model are reported with standard deviations in parentheses.

3.9 Concluding remarks

This paper has investigated the presence of long memory in Realized Volatility (RV) through analysis of the Heterogeneous Autoregressive (HAR) model and fractionally integrated long memory models. We find that the presence of the long memory parameter is often important in addition to the HAR models and that their relative importance seems to vary across the asset process being considered. In several cases the preferred model is a combination of the two approaches. The HAR restricted $ARFIMA$ model, denoted by $RARFIMA$ appears to be a good approximation to the dynamic structure of several RV series.

Time varying parameter versions of the HAR model were also investigated and show the relative importance of different HAR components at different time periods in the sample. In general, the $RARFIMA$ model is preferred to the time varying parameter models on information criteria.

Our results suggest that RV series are quite complex and can involve both HAR components and long memory components. In fact, RV potentially convey a lot of information and are worthy of further research.

Appendix A

Competing models

Starting from the seminal work of Corsi (2009), several different approaches have been proposed to capture the volatility characteristics. All the models are estimated by ordinary least squares with Newey-West (HAC) standard errors correction for serial correlation except the HAR-FIGARCH; these are estimated using quasi-maximum likelihood estimator. Here we briefly survey the main models proposed.

HAR-RV-CJ – Andersen et al.(2007) explicitly decomposed the Realized Volatility in a continuous path and a jump process and employed these estimates of volatility and jumps to model the volatility from today RV. The model is the following:

$$\log RV_t^d = \alpha_0 + \alpha_1 \log C_{t-1}^d + \alpha_2 \log C_{t-1}^w + \alpha_3 \log C_{t-1}^m \\ + \alpha_4 \log(J_{t-1}^d + 1) + \alpha_5 \log(J_{t-1}^w + 1) + \alpha_6 \log(J_{t-1}^m + 1) + \varepsilon_t,$$

where $\log C_{t-1}^w = 1/5 \sum_{i=1}^{5} \log C_{t-i}$ and $\log C_{t-1}^m = 1/22 \sum_{i=1}^{22} \log C_{t-i}$ while $\log(J_{t-1}^w + 1) = \sum_{i=1}^{5} \log(J_{t-i} + 1)$ and $\log(J_{t-1}^m + 1) = \sum_{i=1}^{22} \log(J_{t-i} + 1)$. A forecasting accuracy analysis has shown that mainly the continuous component of volatility has an useful predictive power, while the jump component drag only a residual power.

HAR-RV-CJI – Clements and Liao (2013) amended the previous model, observing that a dimension of jump activity has been neglected. Hence, they proposed to include into the model both jumps magnitude and the probability of jump occurrence, reaching the following expression:

$$\log RV_t^d = \alpha_0 + \alpha_1 \log C_{t-1}^d + \alpha_2 \log C_{t-1}^w + \alpha_3 \log C_{t-1}^m + \alpha_4 \log(J_{t-1}^d + 1) + \alpha_5 \log(J_{t-1}^w + 1) + \alpha_6 \log(J_{t-1}^m + 1) + \alpha_7 \lambda_t + \varepsilon_t,$$

where λ_t is the probability of jump occurrence, computed according equation (14). Their results underline a significant predictive power of both jump components and of continuous part. Further, in contrast with Andersen et al. (2007), HAR-RV-CJI model shows some predictive power also for the magnitude of jumps.

LHAR-RV-CJ – Another extension is to consider the so-called leverage effect. Corsi et al. (2012) proposed to extend the asymmetric behaviour implemented for the volatility to the returns in order to capture the relationship between returns and volatility. In other words the paper accounts for an asymmetric return-volatility dependence at each level, in addition to the classical heterogeneous scheme:

$$\begin{aligned}\log RV_t^d = {} & \alpha_0 + \alpha_1 \log C_{t-1}^d + \alpha_2 \log C_{t-1}^w + \alpha_3 \log C_{t-1}^m \\ & + \alpha_4 \log(J_{t-1}^d + 1) + \alpha_5 \log(J_{t-1}^w + 1) + \alpha_6 \log(J_{t-1}^m + 1) \\ & + \alpha_8 r_t^{d-} + \alpha_9 r_t^{w-} + \alpha_{10} r_t^{m-} + \varepsilon_t,\end{aligned}$$

where $r_t^{m-} = \max(r_t^{(n)}, 0)$ and $r_t^m = 1/n \sum_{i=1}^n r_{t-i}$ with $m \in [d; w; m]$ and $n \in [1, 5, 22]$. This model nests all the other models presented so far, producing some difficulties for the evaluation of their forecasting performances.

HAR-$RS^{+/-}$ – Patton and Shepard (2015), using the exact decomposition of RV into RS^+ ans RS^-, semi variances, proposed another way to consider the contribution of the positive and negative returns to the future volatility, abandoning the heterogeneous scheme adopted by Corsi. They suggested the following model:

$$\log RV_t^d = \alpha_0 + \alpha_1 \log RS_{t-1}^{d+} + \alpha_2 \log RS_{t-1}^{d-} + \alpha_3 \log RV_{t-1}^w + \alpha_4 \log RV_{t-1}^m + \varepsilon_t,$$

where $RS_t^+ = \sum_{i=1}^{M} r_{t,\tau}^2 I\{r_{t,\tau} > 0\}$ and $RS_t^- = \sum_{i=1}^{M} r_{t,\tau}^2 I\{r_{t,\tau} < 0\}$ are respectively the positive and negative semivariance. They found that the two new components, $RS^{+/-}$, are significant and that future volatility is more related with of negative realized semi variance than positive one, and disentangling the effects of these 2 components significantly improves volatility forecasts.

HAR-$RS^{J^{+/-}}$ – Moving from the HAR-$RS^{+/-}$ approach, Patton and Shepard (2015) proposed also to take into account the contribution of signed jump variation, extracted from the different behaviour of the RS: $\Delta J^2 \equiv RS_t^+ - RS_t^-$. Then in order to determine whether the coefficient on positive jump variation differs from that of negative jump one, and thus whether the impact of jumps is driven more by positive or negative jump variation, they extend the model considering: $\Delta J^{2+} = RS_t^+ - RS_t^- I(RS_t^+ - RS_t^- > 0)$ and $\Delta J^{2-} = RS_t^+ - RS_t^- I(RS_t^+ - RS_t^- > 0)$. The model they proposed is the following:

$$\log RV_t^d = \alpha_0 + \alpha_1 \Delta J_{t-1}^{2,d+} + \alpha_2 \Delta J_{2,t-1}^{d-} + \alpha_3 \log BPV_{t-1}^d + \alpha_4 \log RV_{t-1}^w + \alpha_5 \log RV_{t-1}^m + \varepsilon_t,$$

where BPV_t is the Bi-Power Variation proposed by Barndorff-Nielsen and Shephard (2004). They found that the impact of the jump process on volatility depends on the sign of the jump, with negative (positive) jumps leading to higher (lower) future volatility.

HAR-FIGARCH (m, d_u, q) – Louzis et al. (2012) proposed another way to investigate volatility features. The leverage effect is modelled as lagged standardised returns and absolute standardised returns occurring at different time (similar to a more flexible EGARCH model). Whilst, long memory effect is accounted using a Fractionally Integrated GARCH (FIGARCH) model for the conditional heteroschedasticity of the HAR residual. The proposed model is defined as:

$$\begin{aligned}\log(RV_t^d) &= \alpha_0 + \alpha_1 \log RV_{t-1}^d + \alpha_2 \log RV_{t-1}^w + \alpha_3 \log RV_{t-1}^m \\ &+ b_1 z_{t-1}^d + b_2 z_{t-1}^w + b_3 z_{t-1}^m \\ &+ c_1 |z_{t-1}^d| + c_2 |z_{t-1}^w| + c_3 |z_{t-1}^m| + u_t, \end{aligned} \tag{A.1}$$

where $z_t^h = \sum_{i=1}^h r_{t-i+1} / \sum_{i=1}^h \sqrt{\sigma^2_{RV_{t+1-i}}}$ are the daily ($h = d = 1$), weekly ($h = w = 5$) and monthly ($h = m = 22$) standardised returns. Hence, the $\log RV$ will react to past positive and negative returns as follow:

$$\frac{\delta \log RV_t^d}{\delta z_{t-1}^{(\cdot)}} = \begin{cases} b_{(\cdot)} + c_{(\cdot)} & \text{if } z_{t-1}^{(\cdot)} > 0 \\ b_{(\cdot)} - c_{(\cdot)} & \text{if } z_{t-1}^{(\cdot)} < 0. \end{cases}$$

The symmetric behaviour is captured by the coefficients $b_{(.)}$ which are expected to be negative and statistically different from zero. Further, the heteroschedastic variance of residual, $u_t = \sigma_{u,t}\varepsilon_t$ with $\varepsilon\ N(0,1)$, follows a FIGARCH (m, d_u, q) model;

$$\sigma^2_{u,t} = \omega + \beta(L)\sigma^2_t + [1 - \beta(L) - \varphi(L)(1-L)^{d_u}]u_t^2.$$

The fractional differencing parameter d_u captures the long memory behaviour, for values between 0 and 1, and the term $(1-L)^{d_u}$ is an infinite summation that allows an infinite order specification of the FIGARCH, (following Baillie et al. (1996), it is truncated at 1000 lags):

$$\begin{aligned}(1-L) &= \textstyle\sum_{k=0}^{\infty} \frac{\Gamma(d_u+1)}{\Gamma(k+1)\Gamma(d_u-k+1)} L^k \\ &= 1 - d_u L - \tfrac{1}{2} d_u(1-d_u)L^2 \ldots\end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function. The model is estimated with the quasi-maximum estimator, while the optimum lag order for proposed model, that must be estimated before parameters, is founded comparing the AIC and SIC information criteria for different lag structure combinations.

Appendix B

Descriptive Statistics of the volatility measures

In this Section, we report the descriptive statistics for the main volatility and jump measures introduced in Chapter 1. The statistics are presented for each of the 50 stocks under analysis; the details of the stocks are shown in Table 1.2.
Precisely, the following Tables B.1-B.3 provide the following statistics: Square return, (ret^2); Realized Volatility, RV, (Equation 1.3); Bi-Power Variation, BPV, (Equation 1.7), Jump magnitude, J, (Equation 1.18) and Jump intensity Int, (Equation 1.22) where we considered as exogenous variable the lags of the Bi Power Variation: $X_t = BPV_t$[1].
The datasets are obtained from Alpha Trading, (alphatrading.com) and consist in Historical Intra-day ASCII Database, 1-minute tick interpolated prices over more than 17-year period, from 30/01/2002 to 01/05/2017, with the exception of CBS Corp. from 03/01/2006.

[1]Clements and Liao (2013) showed that BPV and other jump robust measures of volatility such as the Threshold Bi-Power Variation , $TBPV$, produces statistically identical results. For this reason, we report here only the results for BPV as exogenous variables. Results for other quantities are available upon request.

Table B.1: Summary statistics for different volatility and jump measures of different assets

		Mean	St. Dev.	Min	Max	Skew	Kurt		Mean	St. Dev.	Min	Max	Skew	Kurt
JPM								ALK						
	ret^2	0.0004	0.0023	0.0000	0.070	17.599	407.97		0.0005	0.0024	0.0000	0.086	19.823	571.58
	RV	0.0363	0.1076	0.0010	2.311	11.463	189.37		0.0599	0.1027	0.0028	3.233	11.972	283.11
	BPV	0.0340	0.1050	0.0008	2.469	12.979	244.34		0.0534	0.0948	0.0011	2.996	11.995	285.39
	J	0.0027	0.0113	0.0000	0.265	11.560	193.09		0.0072	0.0182	0.0000	0.378	11.081	178.30
	Int	0.5720	0.0153	0.4810	0.5976	-1.2271	3.5056		0.7291	0.0474	0.4734	0.8139	-0.6099	1.2419
BLK								EFX						
	ret^2	0.0003	0.0013	0.0000	0.037	13.140	259.33		0.0002	0.0006	0.0000	0.017	12.746	256.11
	RV	0.0310	0.0682	0.0008	1.382	9.207	124.08		0.0171	0.0273	0.0009	0.537	7.523	87.48
	BPV	0.0281	0.0646	0.0001	1.406	10.154	155.37		0.0158	0.0269	0.0008	0.538	8.030	96.71
	J	0.0032	0.0118	0.0000	0.428	19.997	588.87		0.0013	0.0034	0.0000	0.083	8.163	132.90
	Int	0.6610	0.1534	0.1149	0.9354	0.2564	-0.8170		0.5762	0.0681	0.2892	0.7116	-0.2593	-0.4302
BAC								FDX						
	ret^2	0.0007	0.0044	0.0000	0.133	15.403	305.27		0.0002	0.0007	0.0000	0.020	10.418	186.41
	RV	0.0521	0.1839	0.0009	5.586	12.934	277.70		0.0200	0.0317	0.0011	0.708	7.742	100.44
	BPV	0.0480	0.1807	0.0004	6.291	16.067	437.64		0.0188	0.0307	0.0010	0.681	7.905	102.17
	J	0.0046	0.0222	0.0000	0.511	12.711	213.40		0.0012	0.0038	0.0000	0.090	8.384	121.03
	Int	0.6161	0.0000	0.6160	0.6161	-1.0425	6.3254		0.5273	0.0166	0.4326	0.5645	-0.7456	3.8808
AXP								UNP						
	ret^2	0.0003	0.0016	0.0000	0.042	14.070	272.20		0.0002	0.0007	0.0000	0.020	10.171	184.82
	RV	0.0315	0.0939	0.0006	3.193	14.971	398.66		0.0239	0.0462	0.0010	1.100	8.900	128.43
	BPV	0.0293	0.0879	0.0005	3.088	15.599	440.30		0.0223	0.0463	0.0007	1.176	9.990	160.86
	J	0.0024	0.0130	0.0000	0.498	20.320	642.52		0.0018	0.0057	0.0000	0.157	11.083	213.62
	Int	0.4924	0.0491	0.2833	0.5982	-0.4487	-0.0548		0.5971	0.0826	0.2694	0.7708	-0.1198	-0.2547
WFC								KSU						
	ret^2	0.0005	0.0033	0.0000	0.099	19.446	470.13		0.0004	0.0013	0.0000	0.040	12.574	256.57
	RV	0.0395	0.1224	0.0011	2.360	8.250	94.85		0.0413	0.0730	0.0022	2.052	11.414	226.17
	BPV	0.0366	0.1157	0.0008	2.475	8.765	111.76		0.0366	0.0602	0.0019	1.403	8.299	116.52
	J	0.0031	0.0146	0.0000	0.289	11.042	157.63		0.0049	0.0345	0.0000	1.815	43.528	2172.81
	Int	0.4895	0.0780	0.2037	0.6264	-0.5219	-0.4246		0.6841	0.0511	0.4266	0.7974	0.0580	0.6616
BAX								AAPL						
	ret^2	0.0001	0.0008	0.0000	0.044	39.587	2121.10		0.0003	0.0011	0.0000	0.035	13.803	310.21
	RV	0.0155	0.0336	0.0010	1.275	22.345	712.15		0.0379	0.0729	0.0000	2.538	15.893	445.11
	BPV	0.0140	0.0324	0.0009	1.398	26.747	1021.70		0.0304	0.0659	0.0000	2.427	18.826	567.87
	J	0.0013	0.0056	0.0000	0.172	17.958	446.19		0.0078	0.0184	0.0000	0.371	5.136	54.165
	Int	0.5969	0.0782	0.2692	0.7401	-0.3905	-0.3768		0.6537	0.1316	0.2178	0.8909	0.1929	-0.7357
ABT								EBAY						
	ret^2	0.0001	0.0004	0.0000	0.009	10.779	171.61		0.0003	0.0013	0.0000	0.042	15.908	380.29
	RV	0.0146	0.0262	0.0012	0.955	18.476	545.80		0.0304	0.0416	0.0020	0.971	8.292	119.43
	BPV	0.0131	0.0242	0.0008	0.969	20.761	716.65		0.0277	0.0395	0.0014	0.962	8.781	135.54
	J	0.0011	0.0052	0.0000	0.218	27.206	1009.29		0.0028	0.0063	0.0000	0.141	8.382	127.53
	Int	0.6207	0.0542	0.3557	0.7117	-0.8635	0.6233		0.6773	0.0760	0.1144	0.8372	-1.6777	8.8465
JNJ								AMZN						
	ret^2	0.0001	0.0003	0.0000	0.010	17.006	439.09		0.0005	0.0023	0.0000	0.073	19.156	496.63
	RV	0.0089	0.0217	0.0006	0.817	20.366	625.85		0.0380	0.0623	0.0011	1.457	9.876	150.35
	BPV	0.0081	0.0190	0.0005	0.655	18.206	500.13		0.0359	0.0608	0.0011	1.393	9.936	149.45
	J	0.0005	0.0041	0.0000	0.162	28.910	1039.18		0.0027	0.0064	0.0000	0.160	8.634	145.13
	Int	0.3826	0.0821	0.1494	0.6070	0.2477	-0.6370		0.5882	0.0019	0.5774	0.5907	-2.5911	8.7899
MDT								INTC						
	ret^2	0.0001	0.0006	0.0000	0.018	15.631	333.70		0.0002	0.0008	0.0000	0.015	9.515	132.52
	RV	0.0145	0.0370	0.0005	1.318	25.004	810.63		0.0245	0.0387	0.0016	1.165	12.826	289.76
	BPV	0.0132	0.0299	0.0003	1.075	20.685	605.11		0.0225	0.0409	0.0012	1.405	17.060	472.25
	J	0.0011	0.0110	0.0000	0.591	45.453	2366.94		0.0020	0.0044	0.0000	0.098	6.325	83.63
	Int	0.5242	0.0208	0.4084	0.5612	-1.0430	3.0430		0.6616	0.0795	0.1019	0.8147	-1.7842	7.5169
PFE								ADBE						
	ret^2	0.0001	0.0005	0.0000	0.012	12.901	221.58		0.0003	0.0012	0.0000	0.037	12.580	258.86
	RV	0.0175	0.0289	0.0014	0.745	10.164	170.74		0.0276	0.0423	0.0019	1.328	12.048	280.91
	BPV	0.0158	0.0270	0.0008	0.751	11.378	223.24		0.0257	0.0400	0.0017	1.212	11.341	246.75
	J	0.0015	0.0048	0.0000	0.111	11.371	190.78		0.0021	0.0053	0.0000	0.116	7.856	106.82
	Int	0.6519	0.0358	0.4598	0.7139	-1.3284	3.8982		0.6024	0.0004	0.5996	0.6031	-2.1303	9.6823

Note: The Table provides the descriptive statistics for different volatility measures introduced in the main body of this work computed for each stock under analysis. The volatility measures are the following: Square return, (ret^2),T*Bi-Power Variation*, (BPV) ,*Jumps magnitude*, (J), *Jumps intensity*, (int).

Table B.2: Summary statistics for different volatility and jump measures of different assets (cont'd)

		Mean	St. Dev.	Min	Max	Skew	Kurt		Mean	St. Dev.	Min	Max	Skew	Kurt
XOM								VMC						
	ret^2	0.0002	0.0008	0.0000	0.037	25.694	902.36		0.0003	0.0012	0.0000	0.038	13.370	309.12
	RV	0.0165	0.0482	0.0007	2.020	24.578	892.60		0.0380	0.0992	0.0013	4.621	29.208	1283.13
	BPV	0.0156	0.0484	0.0005	2.161	27.771	1115.34		0.0346	0.0616	0.0012	0.968	6.370	54.27
	J	0.0009	0.0043	0.0000	0.152	18.991	541.57		0.0039	0.0618	0.0000	3.653	57.757	3403.70
	Int	0.6610	0.1534	0.1149	0.9354	0.2564	-0.8170		0.6157	0.0000	0.6156	0.6157	-2.1184	10.3231
PXD								WRK						
	ret^2	0.0005	0.0017	0.0000	0.049	12.197	240.36		0.0002	0.0007	0.0000	0.024	16.807	408.90
	RV	0.0487	0.0820	0.0041	2.038	10.045	166.39		0.0186	0.0409	0.0008	1.046	11.110	182.79
	BPV	0.0458	0.0775	0.0033	2.108	10.297	186.06		0.0174	0.0405	0.0006	1.171	12.723	251.12
	J	0.0039	0.0179	0.0000	0.848	31.704	1393.82		0.0013	0.0058	0.0000	0.182	17.902	457.94
	Int	0.6161	0.0000	0.6160	0.6161	-1.0425	6.3254		0.4895	0.0780	0.2037	0.6264	-0.5219	-0.4246
CVX								AVY						
	ret^2	0.0002	0.0009	0.0000	0.041	25.871	955.79		0.0002	0.0009	0.0000	0.021	11.469	180.25
	RV	0.0192	0.0500	0.0009	1.956	20.987	686.93		0.0203	0.0319	0.0010	0.587	6.805	71.06
	BPV	0.0181	0.0495	0.0005	2.028	22.824	806.93		0.0185	0.0288	0.0008	0.478	6.321	58.64
	J	0.0011	0.0041	0.0000	0.110	11.737	214.20		0.0017	0.0076	0.0000	0.344	28.215	1177.05
	Int	0.4924	0.0491	0.2833	0.5982	-0.4487	-0.0548		0.5812	0.0226	0.4492	0.6340	-0.9099	3.3092
APA								APD						
	Ret^2	0.0004	0.0014	0.0000	0.043	15.476	366.09		0.0002	0.0007	0.0000	0.019	12.690	244.20
	RV	0.0369	0.0657	0.0021	1.663	10.790	190.55		0.0192	0.0398	0.0014	0.903	10.219	152.05
	BPV	0.0350	0.0627	0.0010	1.646	10.699	192.07		0.0180	0.0396	0.0010	0.976	11.263	187.02
	J	0.0025	0.0074	0.0000	0.168	10.845	185.59		0.0012	0.0041	0.0000	0.092	9.530	140.16
	Int	0.5657	0.0029	0.5475	0.5717	-1.1845	3.8637		0.5325	0.0451	0.3212	0.6384	-0.5115	0.3825
SLB								EMN						
	Ret^2	0.0003	0.0012	0.0000	0.034	14.156	292.53		0.0003	0.0011	0.0000	0.043	18.099	573.48
	RV	0.0351	0.0617	0.0015	1.670	10.467	185.64		0.0256	0.0411	0.0016	0.847	8.420	109.29
	BPV	0.0332	0.0600	0.0009	1.738	11.539	229.59		0.0234	0.0398	0.0014	0.957	9.484	146.40
	J	0.0023	0.0077	0.0000	0.257	15.049	389.52		0.0022	0.0071	0.0000	0.277	19.928	674.95
	Int	0.5720	0.0153	0.4810	0.5976	-1.2271	3.5056		0.6445	0.0462	0.4093	0.7543	-0.3403	1.0338
DUK								BXP						
	Ret^2	0.0001	0.0005	0.0000	0.022	24.181	886.76		0.0003	0.0018	0.0000	0.060	16.024	365.42
	RV	0.0146	0.0317	0.0005	1.231	20.247	662.93		0.0298	0.0739	0.0008	1.499	7.803	92.90
	BPV	0.0131	0.0274	0.0004	1.000	17.483	518.09		0.0277	0.0704	0.0005	1.565	8.447	113.95
	J	0.0013	0.0061	0.0000	0.231	22.417	711.74		0.0022	0.0087	0.0000	0.132	8.067	80.16
	Int	0.5805	0.1303	0.1718	0.7831	-0.3963	-1.2704		0.5426	0.0923	0.1266	0.7592	0.1561	-0.0187
AEP								FRT						
	Ret^2	0.0001	0.0005	0.0000	0.016	16.798	436.36		0.0003	0.0015	0.0000	0.048	16.727	393.49
	RV	0.0160	0.0360	0.0006	1.209	15.620	397.36		0.0283	0.0734	0.0004	1.556	9.421	136.19
	BPV	0.0145	0.0326	0.0005	1.027	14.390	331.38		0.0257	0.0652	0.0001	1.201	8.350	102.85
	J	0.0013	0.0078	0.0000	0.288	24.119	735.67		0.0027	0.0151	0.0000	0.627	24.002	859.00
	Int	0.5597	0.0657	0.2742	0.7153	-0.0877	-0.2764		0.5978	0.0632	0.3230	0.7457	0.1553	-0.3317
PPL								VTR						
	ret^2	0.0001	0.0006	0.0000	0.020	19.053	492.35		0.0003	0.0017	0.0000	0.042	13.134	220.01
	RV	0.0167	0.0381	0.0011	1.344	16.527	455.77		0.0334	0.0761	0.0011	1.660	8.032	103.63
	BPV	0.0151	0.0350	0.0006	1.221	16.277	441.14		0.0300	0.0683	0.0009	1.407	7.232	80.84
	J	0.0014	0.0061	0.0000	0.159	15.222	296.21		0.0037	0.0140	0.0000	0.342	12.204	204.00
	Int	0.6202	0.0620	0.3350	0.7476	-0.5441	-0.0486		0.6240	0.1080	0.2402	0.8394	0.0117	-0.8143
FE								PSA						
	ret^2	0.0002	0.0007	0.0000	0.028	20.916	665.13		0.0003	0.0015	0.0000	0.051	17.325	431.09
	RV	0.0177	0.0430	0.0007	1.774	22.501	809.57		0.0277	0.0688	0.0011	1.895	11.541	224.06
	BPV	0.0164	0.0403	0.0005	1.626	21.503	748.35		0.0248	0.0570	0.0004	1.495	9.671	165.27
	J	0.0012	0.0049	0.0000	0.148	16.128	378.43		0.0030	0.0256	0.0000	1.354	43.001	2205.50
	Int	0.5845	0.0010	0.5779	0.5863	-1.8747	8.0100		0.5547	0.1118	0.2053	0.8173	0.4501	-0.6123
DE								SPG						
	ret^2	0.0001	0.0003	0.0000	0.010	17.267	477.64		0.0004	0.0021	0.0000	0.055	14.767	283.12
	RV	0.0112	0.0218	0.0006	0.717	14.772	360.11		0.0331	0.0829	0.0002	1.274	6.778	63.48
	BPV	0.0102	0.0210	0.0005	0.741	16.523	454.00		0.0304	0.0744	0.0002	1.155	6.643	61.21
	J	0.0007	0.0031	0.0000	0.130	22.482	836.06		0.0030	0.0178	0.0000	0.635	19.701	549.22
	Int	0.5202	0.0589	0.2718	0.6306	-0.3457	-0.4538		0.5254	0.0697	0.2454	0.6805	0.0722	-0.3988

Note: The Table provides the descriptive statistics for different volatility measures introduced in the main body of this work computed for each stock under analysis. The volatility measures are the following: Square return, (ret^2), *Bi-Power Variation*, (BPV) ,*Jumps magnitude*, (J), *Jumps intensity*, (int).

Table B.3: Summary statistics for different volatility and jump measures of different assets (cont'd)

		Mean	**St. Dev.**	**Min**	**Max**	**Skew**	**Kurt**		**Mean**	**St. Dev.**	**Min**	**Max**	**Skew**	**Kurt**
RL								PM						
	Ret^2	0.0003	0.0014	0.0000	0.049	16.074	408.98		0.0001	0.0006	0.0000	0.017	14.693	299.91
	RV	0.0314	0.0525	0.0013	1.149	7.655	97.78		0.0168	0.0641	0.0009	2.435	25.892	916.26
	BPV	0.0287	0.0494	0.0013	0.924	7.583	89.87		0.0155	0.0607	0.0008	2.350	27.050	981.88
	J	0.0030	0.0111	0.0000	0.315	17.403	393.42		0.0012	0.0060	0.0000	0.111	10.987	148.91
	Int	0.6349	0.0374	0.4400	0.7333	0.0696	1.1924		0.4658	0.0734	0.2386	0.6527	0.5258	-0.2953
DIS								CVS						
	Ret^2	0.0002	0.0008	0.0000	0.021	13.886	273.14		0.0002	0.0009	0.0000	0.040	22.569	748.39
	RV	0.0192	0.0399	0.0010	1.243	14.093	329.93		0.0210	0.0652	0.0010	2.426	25.150	829.95
	BPV	0.0178	0.0377	0.0009	1.180	14.026	326.87		0.0193	0.0564	0.0009	1.940	21.441	629.99
	J	0.0013	0.0045	0.0000	0.086	9.420	127.54		0.0017	0.0115	0.0000	0.486	33.335	1291.93
	Int	0.6519	0.0358	0.4598	0.7139	-1.3284	3.8982		0.5451	0.1019	0.1990	0.7598	-0.4317	-0.7293
TSCO								K						
	Ret^2	0.0003	0.0014	0.0000	0.050	17.724	467.20		0.0001	0.0003	0.0000	0.006	10.261	140.33
	RV	0.0387	0.0498	0.0029	0.963	6.635	84.72		0.0094	0.0166	0.0006	0.593	15.802	458.74
	BPV	0.0342	0.0435	0.0024	0.768	5.608	56.56		0.0086	0.0164	0.0004	0.647	19.575	667.45
	J	0.0048	0.0127	0.0000	0.489	19.439	641.75		0.0005	0.0023	0.0000	0.062	11.607	213.12
	Int	0.6917	0.1296	0.1147	0.9085	-0.2367	-0.5195		0.4745	0.0878	0.1926	0.6695	0.1740	-0.9602
RCL								CL						
	Ret^2	0.0006	0.0023	0.0000	0.054	11.731	194.12		0.0001	0.0004	0.0000	0.013	15.279	329.69
	RV	0.0453	0.0901	0.0014	1.611	6.877	71.82		0.0102	0.0191	0.0007	0.717	17.841	555.66
	BPV	0.0421	0.0866	0.0011	1.684	7.244	80.80		0.0092	0.0175	0.0005	0.671	18.457	595.07
	J	0.0037	0.0110	0.0000	0.197	8.379	102.29		0.0007	0.0039	0.0000	0.158	26.569	955.34
	Int	0.6405	0.0192	0.5243	0.6746	-1.1569	3.0535		0.5126	0.1035	0.1794	0.7061	-0.1540	-1.1758
CBS								KO						
	Ret^2	0.0005	0.0024	0.0000	0.067	14.822	286.68		0.0001	0.0004	0.0000	0.011	16.889	419.23
	RV	0.0451	0.0892	0.0014	1.257	5.499	41.55		0.0104	0.0201	0.0006	0.688	15.280	403.31
	BPV	0.0416	0.0839	0.0010	1.306	5.951	50.72		0.0092	0.0184	0.0005	0.621	15.000	385.38
	J	0.0040	0.0144	0.0000	0.317	10.484	160.53		0.0007	0.0028	0.0000	0.067	11.399	201.56
	Int	0.6211	0.0197	0.5086	0.6587	-1.0200	3.3075		0.5691	0.1174	0.1696	0.7818	-0.2279	-1.0148

Note: The Table provides the descriptive statistics for different volatility measures introduced in the main body of this work computed for each stock under analysis. The volatility measures are the following: Square return, (ret^2), *Bi-Power Variation*, (BPV) ,*Jumps magnitude*, (J), *Jumps intensity*, (int).

Appendix C

In sample - Forecasting Performances Results

In this Section, we report the results for an in-of-sample forecasting exercise, delineated in Chapter 1, using the Root-Mean Square Error as loss function.
Following the procedure adopted in the main body of the text, we employ two different predictive ability test: the Model Confidence Set (MCS) procedure, developed by Hansen et al. (2011) and the Giacomini and White (2006) (GW) test. The former procedure consists of a sequence of tests which permits to construct a set of superior models, where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a certain confidence level. The GW test, instead, helps to identify whether the differences in forecasting performance of competing models are statistically significant. This test is a t-test with heteroskedasticity and autocorrelation consistent (HAC) standard error.

Table C.1: MCS summary results - In sample analysis

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ	LHAR-CJI
Financial									
JPM	-	-	-	-	-	2	-	-	1
BLK	-	-	-	-	-	3	-	2	1
BAC	-	-	-	-	-	1	-	-	2
AXP	-	3	-	-	2	-	-	-	1
WFC	-	4	-	-	3	5	-	2	1
Industrial									
ALK	-	-	-	-	4	2	-	3	1
EFX	-	-	-	-	-	3	-	2	1
FDX	-	-	-	-	-	-	-	-	1
UNP	-	-	-	-	4	2	-	3	1
KSU	-	-	-	-	3	-	-	2	1
Energy									
XOM	-	-	-	-	4	2	3	1	-
PXD	-	-	-	-	4	2	-	3	1
CVX	-	4	-	2	-	3	-	2	1
APA	-	-	-	-	2	-	3	1	-
EMN	-	-	-	-	3	-	-	2	1
Materials									
VMC	-	-	-	6	4	3	5	2	1
WRK	5	6	-	4	3	1	-	-	2
AVY	-	-	-	-	-	4	3	1	2
APD	-	-	-	-	4	2	-	1	3
EMN	-	-	-	4	-	3	-	2	1
Cons. Discr.									
RL	-	-	-	-	4	2	-	3	1
DIS	-	-	-	-	3	2	5	4	1
TSCO	-	-	-	-	-	-	-	-	1
RCL	-	-	-	-	-	-	-	2	1
CBS	-	-	-	-	2	3	-	-	1
Health									
BAX	4	8	-	7	5	1	6	2	3
ABT	3	-	-	7	5	4	6	1	2
JNJ	1	-	-	-	4	-	-	2	3
MDT	2	8	9	7	1	4	6	5	3
PFE	3	7	8	6	4	2	-	5	1
IT									
AAPL	-	-	-	5	3	2	-	4	1
EBAY	-	-	-	-	3	5	4	2	1
AMZN	-	-	-	-	4	2	-	2	1
INTC	-	-	-	-	4	2	5	3	1
ADBE	-	-	-	-	2	1	-	-	3
Utilities									
DUK	-	-	-	-	4	3	-	2	1
AEP	6	7	-	4	-	3	5	2	1
PPL	6	-	-	5	4	1	-	3	2
FE	5	-	-	4	3	1	-	-	2
DE	-	7	-	4	1	5	6	2	3
Real Estate									
BXP	-	5	-	6	4	3	-	2	1
FRT	7	-	-	5	4	2	6	3	1
VTR	-	-	-	3	2	-	-	1	-
PSA	5	-	-	4	3	1	-	2	-
SPG	-	-	-	4	-	2	-	1	3
Cons. Stap.									
PM	8	6	-	7	5	2	4	3	1
CVS	6	7	-	-	4	2	5	1	3
K	-	-	-	3	2	-	-	4	1
CL	-	-	7	4	5	3	-	1	2
KO	-	-	-	4	5	2	6	1	3

Note: The Table provides the summary of the Model Confidence Set (MCS) procedure recently developed by Hansen et al. (2011) for an in sample exercise. Here, we show for each asset whether or not the model in column belongs to the set of the superior forecasting models. the sign (-) means the model does not belong to the set while the numbers indicate the rank produced by the loss function score. The confidence level is set at $\alpha = 0.2$

Table C.2: Giacomini and White test - p values

	$RMSE$	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ
Financial									
HARRV	0.0181	-	-	-	-	-	-	-	-
HAR-RV-J	0.0122	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0151	0.00	0.10	-	-	-	-	-	-
HAR-CJ	0.0135	0.00	0.09	0.00	-	-	-	-	-
$RS^{+/-}$	0.0147	0.00	0.00	0.00	0.00	-	-	-	-
$RS^{J+/-}$	0.0137	0.00	0.00	0.00	0.00	0.05	-	-	-
HAR-CJI	0.0127	0.00	0.00	0.00	0.00	0.00	0.04	-	-
LHAR-CJ	0.0123	0.00	0.00	0.00	0.00	0.00	0.00	0.03	-
LHAR-CJI	0.0081	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Industrial									
HARRV	0.023	-	-	-	-	-	-	-	-
HAR-RV-J	0.0222	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0211	0.00	0.84	-	-	-	-	-	-
HAR-CJ	0.0205	0.00	0.57	0.00	-	-	-	-	-
$RS^{+/-}$	0.0159	0.00	0.83	0.01	0.15	-	-	-	-
$RS^{J+/-}$	0.0147	0.00	0.04	0.00	0.34	0.00	-	-	-
HAR-CJI	0.0175	0.00	0.45	0.00	0.02	0.00	0.38	-	-
LHAR-CJ	0.0155	0.00	0.03	0.00	0.04	0.01	0.18	0.00	-
LHAR-CJI	0.0142	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Energy									
HARRV	0.0193	-	-	-	-	-	-	-	-
HAR-RV-J	0.0283	0.03	-	-	-	-	-	-	-
HAR-FIGARCH	0.0228	0.00	0.56	-	-	-	-	-	-
HAR-CJ	0.0209	0.00	0.43	0.00	-	-	-	-	-
$RS^{+/-}$	0.0167	0.01	0.43	0.03	0.00	-	-	-	-
$RS^{J+/-}$	0.0138	0.01	0.41	0.01	0.02	0.02	-	-	-
HAR-CJI	0.0149	0.01	0.05	0.00	0.02	0.00	0.32	-	-
LHAR-CJ	0.0142	0.00	0.08	0.00	0.01	0.00	0.02	0.05	-
LHAR-CJI	0.0128	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Materials									
HARRV	0.0188	-	-	-	-	-	-	-	-
HAR-RV-J	0.0172	0.02	-	-	-	-	-	-	-
HAR-FIGARCH	0.0199	0.05	0.56	-	-	-	-	-	-
HAR-CJ	0.0199	0.07	0.39	0.00	-	-	-	-	-
$RS^{+/-}$	0.0182	0.00	0.27	0.00	0.00	-	-	-	-
$RS^{J+/-}$	0.0165	0.00	0.02	0.00	0.02	0.03	-	-	-
HAR-CJI	0.0175	0.00	0.16	0.00	0.02	0.02	0.50	-	-
LHAR-CJ	0.0175	0.00	0.00	0.00	0.00	0.02	0.02	0.72	-
LHAR-CJI	0.0165	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Cons. Discr.									
HARRV	0.0081	-	-	-	-	-	-	-	-
HAR-RV-J	0.0081	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0081	0.00	0.71	-	-	-	-	-	-
HAR-CJ	0.0076	0.00	0.21	0.00	-	-	-	-	-
$RS^{+/-}$	0.0062	0.00	0.24	0.00	0.01	-	-	-	-
$RS^{J+/-}$	0.0057	0.00	0.01	0.00	0.01	0.01	-	-	-
HAR-CJI	0.0057	0.00	0.01	0.00	0.00	0.00	0.17	-	-
LHAR-CJ	0.0048	0.02	0.00	0.00	0.00	0.00	0.23	0.03	-
LHAR-CJI	0.0033	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: The Table presents the average *p-values* across the assets belonging to the same sector of GW test, obtained using the in-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark. The datasets are obtained from Alpha Trading, (alphatrading.com) and consist in Historical Intra-day ASCII Database, 1-minute tick interpolated prices over more than 17-year period, from 30/01/2002 to 01/05/2017, with the exception of CBS Corp. from 03/01/2006.

Table C.3: Giacomini and White test - p values (cont'd)

	RMSE	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ
Health									
HARRV	0.0379	-	-	-	-	-	-	-	-
HAR-RV-J	0.0411	0.30	-	-	-	-	-	-	-
HAR-FIGARCH	0.0404	0.18	0.27	-	-	-	-	-	-
HAR-CJ	0.0404	0.13	0.14	0.07	-	-	-	-	-
$RS^{+/-}$	0.0378	0.15	0.02	0.01	0.05	-	-	-	-
$RS^{J^{+/-}}$	0.0385	0.43	0.00	0.00	0.06	0.09	-	-	-
HAR-CJI	0.0388	0.21	0.00	0.00	0.10	0.07	0.28	-	-
LHAR-CJ	0.0385	0.61	0.00	0.00	0.03	0.04	0.71	0.70	-
LHAR-CJI	0.0355	0.05	0.00	0.00	0.00	0.06	0.06	0.05	0.03
IT									
HARRV	0.026	-	-	-	-	-	-	-	-
HAR-RV-J	0.0286	0.43	-	-	-	-	-	-	-
HAR-FIGARCH	0.028	0.00	0.79	-	-	-	-	-	-
HAR-CJ	0.0268	0.00	0.79	0.48	-	-	-	-	-
$RS^{+/-}$	0.0252	0.00	0.41	0.40	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0242	0.03	0.00	0.00	0.00	0.54	-	-	-
HAR-CJI	0.0251	0.00	0.00	0.25	0.26	0.20	0.29	-	-
LHAR-CJ	0.0252	0.00	0.02	0.00	0.33	0.16	0.18	0.23	-
LHAR-CJI	0.023	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Utilities									
HARRV	0.0136	-	-	-	-	-	-	-	-
HAR-RV-J	0.1217	0.05	-	-	-	-	-	-	-
HAR-FIGARCH	0.1038	0.00	0.38	-	-	-	-	-	-
HAR-CJ	0.0114	0.00	0.30	0.42	-	-	-	-	-
$RS^{+/-}$	0.0103	0.00	0.26	0.33	0.80	-	-	-	-
$RS^{J^{+/-}}$	0.0086	0.00	0.02	0.00	0.00	0.00	-	-	-
HAR-CJI	0.0086	0.00	0.02	0.00	0.00	0.00	0.43	-	-
LHAR-CJ	0.0072	0.00	0.03	0.00	0.00	0.00	0.06	0.07	-
LHAR-CJI	0.0067	0.00	0.04	0.00	0.00	0.00	0.38	0.04	0.02
Real Estate									
HARRV	0.0145	-	-	-	-	-	-	-	-
HAR-RV-J	0.0165	0.23	-	-	-	-	-	-	-
HAR-FIGARCH	0.0137	0.23	0.28	-	-	-	-	-	-
HAR-CJ	0.0138	0.22	0.25	0.78	-	-	-	-	-
$RS^{+/-}$	0.013	0.35	0.20	0.63	0.06	-	-	-	-
$RS^{J^{+/-}}$	0.0134	0.14	0.16	0.00	0.63	0.78	-	-	-
HAR-CJI	0.0131	0.10	0.17	0.00	0.03	0.75	0.89	-	-
LHAR-CJ	0.0116	0.04	0.09	0.00	0.00	0.00	0.00	0.00	-
LHAR-CJI	0.0113	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.06
Cons. Stap.									
HARRV	0.0429	-	-	-	-	-	-	-	-
HAR-RV-J	0.0419	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0402	0.00	0.28	-	-	-	-	-	-
HAR-CJ	0.039	0.02	0.25	0.35	-	-	-	-	-
$RS^{+/-}$	0.0388	0.02	0.09	0.02	0.38	-	-	-	-
$RS^{J^{+/-}}$	0.0384	0.01	0.00	0.00	0.33	0.32	-	-	-
HAR-CJI	0.0397	0.01	0.05	0.04	0.30	0.50	0.74	-	-
LHAR-CJ	0.036	0.01	0.00	0.03	0.03	0.28	0.59	0.26	-
LHAR-CJI	0.0336	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06

Note: The Table presents the the average *p-values* across the assets belonging to the same sector of GW test, obtained using the in-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark. The datasets are obtained from Alpha Trading, (alphatrading.com) and consist in Historical Intra-day ASCII Database, 1-minute tick interpolated prices over more than 17-year period, from 30/01/2002 to 01/05/2017, with the exception of CBS Corp. from 03/01/2006.

Appendix D

Forecasting Performances Results with QLIKE

In this Section, we report the results for our out-of-sample forecasting exercise, delineated in Chapter 1, employing the Gaussian Quasi Likelihood ($QLIKE$) loss function. Let define the actual values σ_t, and the forecasts h_t with $t = \tau + 1, \ldots, T$. Hence, the forecast error is: $e_{H+j} = \sigma_{H+j} - h_{H+j}$, $j = H + 1, \ldots, T - 1$. Based on this quantity the $QLIKE$ loss function is:

$$QLIKE = T^{-1} \sum_{t=\tau+1}^{T} (\log(h_t^2) - \sigma_t^2 h_t^{-2})..$$

Such function has been included, since it has been proved to be robust to noise in the proxy for volatility by Patton (2011) and Patton and Sheppard (2009).
Following the procedure adopted in the main body of the text, we use two different predictive ability test: the Model Confidence Set procedure (MCS), developed by Hansen et al. (2011) and the Giacomini and White (2006) test (GW). The former procedure consists on a sequence of tests which permits to construct a set of superior models, where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a certain confidence level. The GW test, instead, helps to identify whether the differences in forecasting performance of competing models are statistically significant.

Table D.1: MCS summary results - QLIKE loss function

	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ	LHAR-CJI
Financial									
JPM	-	-	-	-	-	-	-	-	1
BLK	-	-	-	-	-	-	-	-	1
BAC	-	-	-	-	-	-	-	-	1
AXP	-	-	-	-	-	-	-	-	1
WFC	-	-	-	-	-	2	-	-	1
Industrial									
ALK	-	-	-	-	-	-	-	-	1
EFX	-	-	-	-	-	2	-	-	1
FDX	-	-	-	-	-	-	2	-	1
UNP	-	-	-	-	3	-	2	-	1
KSU	-	-	-	-	3	-	-	2	1
Energy									
XOM	-	-	-	-	-	2	-	3	1
PXD	-	-	-	-	4	3	-	2	1
CVX	-	-	-	4	-	2	-	3	1
APA	-	-	-	-	2	1	-	-	-
EMN	3	-	-	-	-	2	-	1	-
Materials									
VMC	-	-	-	-	4	3	5	2	1
WRK	4	-	-	-	2	1	-	-	3
AVY	-	5	-	-	-	-	2	1	3
APD	-	-	-	-	-	3	-	1	2
EMN	-	-	-	-	-	3	-	2	1
Cons. Discr.									
RL	-	-	-	-	4	2	-	3	1
DIS	-	-	-	-	4	3	-	2	1
TSCO	-	-	-	-	-	-	-	-	1
RCL	-	-	-	-	-	-	-	-	1
CBS	-	-	-	-	2	-	-	-	1
Health									
BAX	7	8	-	5	4	1	2	6	3
ABT	3	-	-	-	-	2	-	4	1
JNJ	1	-	-	-	-	-	-	-	2
MDT	3	-	-	-	2	1	4	-	-
PFE	1	6	-	4	5	3	-	-	2
IT									
AAPL	-	-	-	4	-	2	-	3	1
EBAY	-	-	-	-	-	2	-	3	1
AMZN	-	-	-	-	-	-	-	-	1
INTC	-	-	-	-	-	-	-	2	1
ADBE	-	-	-	-	-	3	-	1	2
Utilities									
DUK	-	-	-	-	-	3	-	2	1
AEP	6	-	5	-	-	1	4	3	2
PPL	-	6	5	4	1	3	-	-	2
FE	5	-	-	1	4	2	-	-	3
DE	-	-	-	6	2	4	3	5	1
Real Estate									
BXP	3	-	-	-	2	-	-	1	-
FRT	7	-	-	4	6	2	5	3	1
VTR	-	-	-	-	4	3	-	2	1
PSA	5	-	-	4	3	-	-	3	1
SPG	-	-	-	4	5	1	-	2	3
Cons. Stap.									
PM	-	6	-	5	4	3	-	1	2
CVS	6	7	9	8	5	3	4	1	2
K	6	-	-	4	5	3	-	2	1
CL	-	-	6	5	3	2	4	-	1
KO	8	7	-	6	5	3	4	1	2

Note: The table provides the summary of the Model Confidence Set (MCS) procedure developed by Hansen et al. (2011). Here, we show for each asset whether or not the model in column belongs to the set of the superior forecasting models. The sign (-) means the model does not belong to the set while the numbers indicate the rank produced by the loss function score. The confidence level is set at $\alpha = 0.2$.

Table D.2: Giacomini and White test - p values

	$QLIKE$	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J^{+/-}}$	HAR-CJI	LHAR-CJ
Financial									
HARRV	0.0232	-	-	-	-	-	-	-	-
HAR-RV-J	0.0156	0.03	-	-	-	-	-	-	-
HAR-FIGARCH	0.0194	0.00	0.39	-	-	-	-	-	-
HAR-CJ	0.0173	0.00	0.38	0.00	-	-	-	-	-
$RS^{+/-}$	0.0189	0.00	0.00	0.00	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0176	0.00	0.01	0.00	0.00	0.20	-	-	-
HAR-CJI	0.0163	0.00	0.00	0.00	0.00	0.00	0.18	-	-
LHAR-CJ	0.0157	0.00	0.01	0.00	0.00	0.00	0.00	0.13	-
LHAR-CJI	0.0104	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Industrial									
HARRV	0.0206	-	-	-	-	-	-	-	-
HAR-RV-J	0.0199	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0188	0.00	0.72	-	-	-	-	-	-
HAR-CJ	0.0183	0.00	0.49	0.00	-	-	-	-	-
$RS^{+/-}$	0.0142	0.00	0.71	0.01	0.25	-	-	-	-
$RS^{J^{+/-}}$	0.0131	0.00	0.03	0.00	0.54	0.00	-	-	-
HAR-CJI	0.0156	0.00	0.39	0.00	0.03	0.00	0.47	-	-
LHAR-CJ	0.0139	0.00	0.03	0.00	0.07	0.01	0.22	0.00	-
LHAR-CJI	0.0127	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Energy									
HARRV	0.0159	-	-	-	-	-	-	-	-
HAR-RV-J	0.0234	0.02	-	-	-	-	-	-	-
HAR-FIGARCH	0.0188	0.00	0.72	-	-	-	-	-	-
HAR-CJ	0.0173	0.00	0.55	0.00	-	-	-	-	-
$RS^{+/-}$	0.0147	0.00	0.55	0.02	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0121	0.00	0.69	0.02	0.04	0.03	-	-	-
HAR-CJI	0.013	0.01	0.08	0.00	0.03	0.01	0.58	-	-
LHAR-CJ	0.0125	0.00	0.13	0.00	0.02	0.00	0.04	0.04	-
LHAR-CJI	0.0112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Materials									
HARRV	0.0171	-	-	-	-	-	-	-	-
HAR-RV-J	0.0156	0.02	-	-	-	-	-	-	-
HAR-FIGARCH	0.0181	0.06	0.76	-	-	-	-	-	-
HAR-CJ	0.0181	0.13	0.71	0.00	-	-	-	-	-
$RS^{+/-}$	0.0164	0.00	0.63	0.00	0.00	-	-	-	-
$RS^{J^{+/-}}$	0.0149	0.00	0.04	0.00	0.05	0.05	-	-	-
HAR-CJI	0.0158	0.00	0.37	0.00	0.04	0.03	0.80	-	-
LHAR-CJ	0.0158	0.00	0.01	0.00	0.01	0.04	0.05	0.02	-
LHAR-CJI	0.0149	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
Cons. Discr.									
HARRV	0.0108	-	-	-	-	-	-	-	-
HAR-RV-J	0.0108	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0108	0.00	0.56	-	-	-	-	-	-
HAR-CJ	0.0107	0.00	0.49	0.00	-	-	-	-	-
$RS^{+/-}$	0.0087	0.00	0.56	0.00	0.03	-	-	-	-
$RS^{J^{+/-}}$	0.0079	0.00	0.02	0.00	0.02	0.02	-	-	-
HAR-CJI	0.0079	0.00	0.02	0.00	0.00	0.00	0.27	-	-
LHAR-CJ	0.0067	0.02	0.00	0.00	0.00	0.00	0.31	0.04	-
LHAR-CJI	0.0047	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: The table presents the average *p-values* across the assets belonging to the same sector of GW test, obtained using the in-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark. The datasets are obtained from Alpha Trading, (alphatrading.com) and consist in Historical Intra-day ASCII Database, 1-minute tick interpolated prices over more than 17-year period, from 30/01/2002 to 01/05/2017, with the exception of CBS Corp. from 03/01/2006.

Table D.3: Giacomini and White test - p values (cont'd)

	QLIKE	HAR-RV	HAR-RV-J	HAR-FIGARCH	HAR-CJ	HAR-$RS^{+/-}$	HAR-$RS^{J+/-}$	HAR-CJI	LHAR-CJ
Health									
HARRV	0.0337	-	-	-	-	-	-	-	-
HAR-RV-J	0.0366	0.34	-	-	-	-	-	-	-
HAR-FIGARCH	0.0359	0.20	0.39	-	-	-	-	-	-
HAR-CJ	0.0359	0.15	0.20	0.10	-	-	-	-	-
$RS^{+/-}$	0.0349	0.17	0.03	0.01	0.07	-	-	-	-
$RS^{J+/-}$	0.0355	0.49	0.00	0.00	0.02	0.03	-	-	-
HAR-CJI	0.0355	0.23	0.00	0.00	0.03	0.02	0.36	-	-
LHAR-CJ	0.0355	0.69	0.00	0.00	0.03	0.04	0.33	0.02	-
LHAR-CJI	0.0328	0.28	0.00	0.00	0.00	0.07	0.08	0.07	0.03
IT									
HARRV	0.0232	-	-	-	-	-	-	-	-
HAR-RV-J	0.0236	0.39	-	-	-	-	-	-	-
HAR-FIGARCH	0.0222	0.00	0.71	-	-	-	-	-	-
HAR-CJ	0.0221	0.00	0.70	0.43	-	-	-	-	-
$RS^{+/-}$	0.0195	0.00	0.33	0.36	0.00	-	-	-	-
$RS^{J+/-}$	0.0195	0.02	0.00	0.00	0.00	0.43	-	-	-
HAR-CJI	0.0203	0.00	0.00	0.22	0.21	0.16	0.24	-	-
LHAR-CJ	0.0203	0.00	0.00	0.00	0.26	0.13	0.14	0.17	-
LHAR-CJI	0.0186	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Utilities									
HARRV	0.0098	-	-	-	-	-	-	-	-
HAR-RV-J	0.088	0.04	-	-	-	-	-	-	-
HAR-FIGARCH	0.0878	0.00	0.41	-	-	-	-	-	-
HAR-CJ	0.0097	0.00	0.33	0.35	-	-	-	-	-
$RS^{+/-}$	0.0087	0.00	0.28	0.28	0.68	-	-	-	-
$RS^{J+/-}$	0.0079	0.00	0.02	0.00	0.00	0.00	-	-	-
HAR-CJI	0.0079	0.00	0.02	0.00	0.00	0.00	0.39	-	-
LHAR-CJ	0.0067	0.00	0.04	0.00	0.00	0.00	0.36	0.42	-
LHAR-CJI	0.0066	0.00	0.04	0.00	0.00	0.00	0.35	0.04	0.02
Real Estate									
HARRV	0.0186	-	-	-	-	-	-	-	-
HAR-RV-J	0.0212	0.30	-	-	-	-	-	-	-
HAR-FIGARCH	0.0177	0.29	0.37	-	-	-	-	-	-
HAR-CJ	0.0178	0.29	0.32	0.01	-	-	-	-	-
$RS^{+/-}$	0.0167	0.45	0.26	0.01	0.00	-	-	-	-
$RS^{J+/-}$	0.0172	0.19	0.21	0.00	0.00	0.00	-	-	-
HAR-CJI	0.0176	0.13	0.23	0.00	0.03	0.00	0.00	-	-
LHAR-CJ	0.0156	0.06	0.12	0.00	0.00	0.00	0.19	0.00	-
LHAR-CJI	0.0152	0.06	0.01	0.01	0.00	0.00	0.00	0.00	0.08
Cons. Stap.									
HARRV	0.0315	-	-	-	-	-	-	-	-
HAR-RV-J	0.0311	0.00	-	-	-	-	-	-	-
HAR-FIGARCH	0.0319	0.00	0.24	-	-	-	-	-	-
HAR-CJ	0.0309	0.02	0.21	0.30	-	-	-	-	-
$RS^{+/-}$	0.0307	0.02	0.08	0.00	0.08	-	-	-	-
$RS^{J+/-}$	0.0305	0.01	0.00	0.00	0.07	0.07	-	-	-
HAR-CJI	0.0315	0.01	0.00	0.00	0.06	0.31	0.56	-	-
LHAR-CJ	0.031	0.01	0.00	0.00	0.07	0.06	0.53	0.06	-
LHAR-CJI	0.029	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05

Note: The table presents the average *p-values* across the assets belonging to the same sector of GW test, obtained using the in-of-sample RMSE. The benchmark model in column is tested against its competitors, model k, with $k = 1, ..., 4$. Under the null hypo-thesis the two model have the same accuracy while under the alternative the method in the row performs better than the benchmark. The datasets are obtained from Alpha Trading, (alphatrading.com) and consist in Historical Intra-day ASCII Database, 1-minute tick interpolated prices over more than 17-year period, from 30/01/2002 to 01/05/2017, with the exception of CBS Corp. from 03/01/2006.

Appendix E

Kernel weighted regression

In order to account for time variation in the coefficients our models, we implement a non parametric kernel approach that has the main advantage of requiring requiring minimal theoretical restriction on the functional form. Specifically, we extend the work of Giraitis, Kapetanios and Yates (2014) work on autoregressive processes to a kernel smoothing regression. Giraitis, Kapetanios and Yates (2014) consider the $AR(1)$ process

$$y_t = \phi_{t-1} y_{t-1} + u_t, \tag{E.1}$$

where u_t is iid $(0, \sigma_u^2)$ and there is some initialization of the process y_0 whereas ϕ_{t-1} is a random coefficient and $u_t|\Omega_{t-1} = 0$ and $\phi_t|\Omega_{t-1} = \phi_t$. The stability of the model depends on the TVP nature of the AR parameters satisfying various smoothness conditions. Giraitis, Kapetanios and Yates (2014) model the TVP parameter, denoted by ϕ_t, for an $AR(1)$ as a rescaled random walk, where $\{a_t\}$ is a non stationary process which defines the random drift, and $-1 < \phi < 1$. In this context ϕ_t is a standardized version of a_t so that

$$\phi_t = \phi \frac{a_t}{\max_{0 \leq k \leq t} |a_k|}, \ldots, t > 0, \tag{E.2}$$

where the stochastic process a_t is assumed to be a drift-less random walk, so that $a_t = a_{t-1} + w_t$ and where w_t is a stationary process with zero mean. Also, $\phi \in (0, 1)$ and ϕ_{t-1} is bounded away from the boundary points of -1 and 1. The above framework can be extended to the time varying $AR(p)$ model

$$y_t = \sum_{i=1}^{p} \phi_{t-1,i} y_{t-i} + u_t$$

and can be used with the boundary conditions

$$\phi_{t,i} = \phi \frac{a_{t,i}}{\max_{0 \le k \le t} |a_{k,i}|}, \ldots, t > 1, \tag{E.3}$$

where $0 < \phi < 1$ and each $a_{t,i}$ are independent versions of the a_t process defined above. Under these assumptions the maximum absolute eigenvalues of the matrix

$$A_t = \begin{bmatrix} \phi_{t,1} & \phi_{t,2} & \ldots & \ldots & \ldots & \phi_{t,p} \\ 1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & \ldots & 0 \\ \ldots & 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & \ldots & 0 & 1 & 0 \end{bmatrix}$$

are bounded above by unity for all t. Giraitis, Kapetanios and Yates (2014) show that the coefficient process $\{\phi_t; t = 1, ..., T\}$ converges in distribution as T increases to the limit

$$\{\phi_t; 0 \le \tau \le 1\} \to_D \{\phi \tilde{W}_\tau; 0 \le \tau \le 1\},$$

where $\tilde{W}_{(.)}$ is the standard Brownian motion. The approach for estimating the time varying parameter, ϕ_t is to use the moving window estimator for the $AR(1)$ RC model

$$\widehat{\phi}_t = \frac{\sum_{t=1}^{H} K\left(\frac{t-k}{H}\right) y_t y_{t-1}}{\sum_{t=1}^{H} K\left(\frac{t-k}{H}\right) y_{t-1}^2}, \tag{E.4}$$

where $K\left(\frac{t-k}{H}\right)$ is a kernel and continuously bounded function, such as the Epanechnikov kernel with finite support, or the familiar Gaussian kernel with infinite support. Generalising this estimation method, a regression can be expressed as

$$y_t = x_t' \beta_t + u_t, \tag{E.5}$$

with $\beta_t = (\beta_{1,t}, \beta_{2,t}, ...\beta_{k,t})$, it is assumed that each $\beta_{j,t}$ follows a bounded random walk. x'_t is the matrix $(m \times T)$ containing the time series of the factors. Therefore, the kernel weighted regression estimator for $\beta_{j,t}$ is

$$\hat{\beta}_{j,t} = \left(\sum_{j=1}^{T} w_{jt} x_j x'_j \right)^{-1} \left(\sum_{j=1}^{T} w_{jt} x_j y_j \right), \tag{E.6}$$

where $w_{jt} = K\left(\frac{t-k}{H}\right)$. The authors prove that if the bandwidth is $o_p\left(T^h\right)$ with $h = 1/2$, and given homoskedasticity of the error process, then

$$Var\left(\hat{\beta}_t\right) = \hat{\sigma}_u^2 \left(\sum_{j=1}^{T} w_{jt} x_j x'_j \right)^{-1} \sum_{j=1}^{T} w_{jt}^2 x_j x'_j \left(\sum_{j=1}^{T} w_{jt} x_j x'_j \right)^{-1}, \tag{E.7}$$

where $\hat{\sigma}_u^2 = \frac{1}{T} \sum_{i=1}^{T} \left(y_t - x'_t \beta_t\right)^2$. While if u_t is heteroskedastic then the covariance matrix of the TVP parameter estimates is given by

$$Var\left(\hat{\beta}_{j,t}\right) = \left(\sum_{j=1}^{T} w_{jt} x_j x'_j \right)^{-1} \left(\sum_{j=1}^{T} w_{jt}^2 x_j x'_j \hat{u}_t^2 \right) \left(\sum_{j=1}^{T} w_{jt} x_j x'_j \right)^{-1}, \tag{E.8}$$

which can be used for inference. One appealing characteristic of this approach is that it nests rolling window estimates of the regression betas and is equivalent to kernel smoothing estimators using a uniform one-sided kernel instead of a Gaussian two-sided kernel. A key role is played by the decision about the selection of the bandwidth and for a given kernel function, $K\left(\frac{t-k}{H}\right)$, the bandwidth, H, represents the degree of smoothness of the estimates. Giraitis, Kapetanios and Yates (2014) proved that a bandwidth of $H = T^h$, with $h = 0.5$, provides an estimator with desirable properties such as consistency and asymptotic normality and in addition provides valid standard errors.

Another appealing characteristics of such approach is that it nests, as a special case, rolling window estimates of betas (for example, Chen, Roll, and Ross, 1986; Ferson and Harvey, 1991; Petkova and Zhang, 2005; among many others). Rolling beta estimates are equivalent to kernel smoothing estimators obtained using a uniform one-sided kernel instead of a Gaussian two-sided kernel, and it has been proved that the order of the smoothing bias of the estimator for the betas and the price of risk parameters is larger for one-sided kernels.

In the kernel estimation approach a key role is played by the selection of the bandwidth. For a given kernel function, $K\left(\frac{t-k}{H}\right)$, the bandwidth, H, represents and controls the degree of smoothness of the estimates. In other terms, if the bandwidth is small, the estimates will be under−smoothed, with high variability, otherwise if the value of H is big, the resulting estimators will be over−smoothed and further from the real function. Different approaches have been proposed to handle the choice of the bandwidth. Ang and Kristensen (2012) suggest to optimise the choice of the bandwidth for conditional and long estimates in order to reduce any finite-sample biases and variances. Giraitis et al. (2014), instead, proved that if the bandwidth is $H = T^h$, with $h = 0.5$, the estimator has desirable properties such as consistency and asymptotic normality and in addition provides valid standard errors.

Appendix F

Robustness checks

Table F.1: Percentage reduction of RMSEs for different bandwidth parameters intervals

	25 Portfolio			*55 Portfolio*			*200 Stocks*		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
$h \in [0.05; 0.9]$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
$h \in [0.05; 0.6]$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.412	-0.413	-1.743
Polling	-2.086	-1.640	-3.800	-1.921	-0.963	-3.585	-1.499	-0.751	-2.796
Average	-1.994	-1.863	-3.625	-2.428	-1.786	-3.446	-1.894	-1.393	-2.688
Specific	-6.970	-6.638	-8.256	-7.374	-6.522	-8.315	-5.752	-5.087	-6.486
$h \in [0.35; 0.9]$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.412	-0.413	-1.743
Polling	0.014	0.087	-1.758	0.028	0.047	-1.714	0.022	0.036	-1.337
Average	0.030	0.109	-1.754	0.075	0.100	-1.671	0.058	0.078	-1.304
Specific	-0.891	-0.766	-2.306	-0.654	-0.517	-2.210	-0.510	-0.403	-1.724
$h \in [0.25; 0.75]$									
h=0.5	-0.522	-0.483	-1.056	-0.528	-0.529	-1.119	-0.412	-0.413	-0.872
Polling	-0.382	-0.245	-0.766	-0.279	-0.200	-0.740	-0.218	-0.156	-0.577
Average	-0.339	-0.288	-0.738	-0.155	-0.087	-0.714	-0.121	-0.068	-0.557
Specific	-2.014	-1.855	-2.029	-1.900	-1.533	-1.842	-1.482	-1.196	-1.436

Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise comparing 4 different intervals for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Table F.2: Percentage reduction of RMSEs for different sample

	25 Portfolio			*55 Portfolio*			*200 Stocks*		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
08/1973 - 05/2016									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
08/1973 - 08/2007									
h=0.5	0.334	0.719	-1.033	-0.947	-0.608	-2.263	-0.794	-0.510	-1.896
Polling	0.401	0.956	-0.959	-0.717	-0.238	-2.076	-0.601	-0.199	-1.740
Average	0.471	0.921	-0.913	-0.648	-0.151	-1.926	-0.543	-0.127	-1.614
Specific	-2.688	-1.904	-3.925	-3.233	-2.369	-4.696	-2.709	-1.986	-3.935

Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise comparing different sub samples identified around the Global financial crisis: 08/1973 - 08/2007. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Table F.3: Percentage reduction of RMSEs for different penalization parameters in LASSO context

	25 Portfolio			*55 Portfolio*			*200 Stocks*		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
No Lasso									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
$\lambda = 0.0001$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.328	-0.328	-1.386
Polling	-0.556	-0.369	-2.496	-0.311	-0.142	-1.983	-0.193	-0.088	-1.229
Average	-1.294	-1.352	-4.975	-0.158	-0.019	-1.810	-0.098	-0.012	-1.122
Specific	-6.821	-6.271	-10.188	-5.229	-4.625	-6.066	-3.242	-2.868	-3.761
$\lambda = 0.00005$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.328	-0.328	-1.386
Polling	-0.469	-0.229	-2.196	-0.311	-0.142	-1.983	-0.193	-0.088	-1.229
Average	-0.514	-0.410	-2.202	-0.134	-0.026	-1.806	-0.083	-0.016	-1.120
Specific	-6.417	-5.661	-6.676	-5.197	-4.617	-6.015	-3.222	-2.862	-3.729

Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise when we consider the LASSO procedure inside our mechanism for the identification of the optimal bandwidth. Here, we report the results for 2 values of the penalty function, λ: 0.0001 and 0.00005. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Table F.4: Percentage reduction of RMSEs for different sample size of the trading period

	25 Portfolio			***55 Portfolio***			***200 Stocks***		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
$T = 60$									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
$T = 180$									
h=0.5	1.629	2.964	-0.490	2.641	4.327	0.435	2.213	3.626	0.365
Polling	1.724	3.522	-0.430	3.055	5.114	0.809	2.560	4.286	0.678
Average	1.986	3.558	-0.160	3.613	5.488	1.377	3.027	4.599	1.154
Specific	-3.454	-1.310	-5.815	-0.952	1.820	-3.835	-0.798	1.525	-3.214
$T = 120$									
h=0.5	0.534	1.152	-1.654	-1.517	-0.974	-3.623	-1.271	-0.816	-3.036
Polling	0.642	1.531	-1.535	-1.148	-0.381	-3.324	-0.962	-0.319	-2.786
Average	0.753	1.474	-1.461	-1.037	-0.242	-3.084	-0.869	-0.203	-2.585
Specific	-4.303	-3.049	-6.284	-5.176	-3.794	-7.518	-4.338	-3.179	-6.300

Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise comparing 3 different trading period, T, for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Table F.5: Percentage reduction of RMSEs for different asset pricing models

	25 Portfolio			***55 Portfolio***			***200 Stocks***		
	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}	$h_{w=12}$	$h_{w=24}$	h_{Kern}
3 Factors									
h=0.5	-0.522	-0.483	-2.302	-0.528	-0.529	-2.235	-0.535	-0.664	-1.902
Polling	-0.396	-0.168	-2.140	-0.219	-0.196	-1.943	0.033	-0.014	-1.382
Average	-0.424	-0.444	-2.048	-0.147	0.019	-1.871	0.102	-0.107	-1.413
Specific	-5.800	-5.383	-6.738	-5.349	-4.444	-6.339	-2.393	-2.119	-3.310
MOM Factor									
h=0.5	-0.569	-0.526	-2.508	-0.576	-0.577	-2.435	-0.357	-0.358	-1.510
Polling	-0.606	-0.402	-2.720	-0.339	-0.155	-2.160	-0.210	-0.096	-1.339
Average	-1.409	-1.473	-5.420	-0.172	-0.021	-1.972	-0.106	-0.013	-1.223
Specific	-7.432	-6.832	-11.099	-5.697	-5.039	-6.608	-3.532	-3.124	-4.097
5 Factors									
h=0.5	-0.655	-0.605	-2.889	-0.663	-0.664	-2.805	-0.411	-0.412	-1.739
Polling	-0.589	-0.287	-2.756	-0.390	-0.179	-2.488	-0.242	-0.111	-1.543
Average	-0.645	-0.514	-2.763	-0.168	-0.032	-2.266	-0.104	-0.020	-1.405
Specific	-8.053	-7.103	-8.378	-6.522	-5.793	-7.548	-4.044	-3.592	-4.680

Note: The table provides the $RMSE$ for the out of sample one step ahead forecasting exercise comparing 3 different asset pricing models: 3 factors Fama and French (1992), momentum factor by Carhart (1997) and the 5 factors model by Fama and French (2015) for the identification of the optimal bandwidth parameter. The results are expressed as a deviation from the $RMSE$ produced by the benchmark model, $FMcB$.

Bibliography

[1] Abadir, K.M., Distaso, W., Giraitis, L. (2007). Nonstationary extended Local Whittle estimation. *Journal of Econometrics*, 141, 1353–1384.

[2] Adrian, T., Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, 16(4), 537–556.

[3] Adrian, T., Crump, R.K. and Moench, E. (2015). Regression based estimation of dynamic asset pricing models. *Journal of Financial Economics*, 118, 211–244.

[4] Aït-Sahalia, Y., Mykland, P., Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18, 351–416.

[5] Andersen, T. G., Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.

[6] Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and Correlation Forecasting, in the Handbook of Economic Forecasting, G. Elliott, CWJ Granger and A. Timmermann eds. *North Holland Press, Amsterdam. BA Benet (1992). Hedge period length and ex ante futures hedging effectiveness: the case of foreign exchange risk cross hedges. The Journal of Futures Markets*, 12, 163–175.

[7] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2000). Exchange rate returns standardized by realized volatility are (nearly) Gaussian, *Multinational Finance Journal*, 4, 159–179.

[8] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, 42–55.

[9] Andersen, T.G., Bollerslev T., Diebold, F.X., Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*, 71, 579–625.

[10] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, G. (2006). Realized beta: Persistence and predictability. *Econometric Analysis of Financial and Economic Time Series*, Emerald Group Publishing Limited, 1–39.

[11] Andersen, T. G., Bollerslev, T., Meddahi, N. (2004). Analytic evaluation of volatility forecast. *International Economic Review*, 45, 1079–1110.

[12] Andersen, T.G., T. Bollerslev, N. Meddahi (2007). Correcting the errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities. *Econometrica*, 73, 279–296.

[13] Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics*, 17, 579–600.

[14] Andreou, E. and Ghysels E. (2006). Monitoring disruptions in financial markets, *Journal of Econometrics*, 135, 77–124.

[15] Ang, A., Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1), 132–156.

[16] Bai, J. (1997). Estimating multiple breaks one at a time, *Econometric Theory*, 13, 315–352

[17] Bai, J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, 66, 47–78.

[18] Bai, J. and Perron P. (2003), Computation and analysis of multiple structural change models, *Journal of Applied Econometrics*, 18(1), 1–22.

[19] Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73, 5–59.

[20] Baillie, R.T., Bollerslev, T., Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.

[21] Baillie, R.T., Kapetanios, G. (2008). Nonlinear models for strongly dependent processes with financial applications. *Journal of Econometrics*, 147, 6–71.

[22] Baillie, R.T., Kapetanios, G. (2013). Estimation and inference for impulse response functions from univariate strongly persistent processes. *Econometrics Journal*, 16, 373–399.

[23] Bali, T. G., Engle, R. F. (2010). The inter-temporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, 57(4), 377–390.

[24] Bandi, F., Russell, J. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies 75*, 339–369.

[25] Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76, 1481–1536.

[26] Barndorff-Nielsen, O.E., Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, 64, 253–280.

[27] Barndorff-Nielsen, O.E., Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2, 1–48.

[28] Barndorff-Nielsen, O., E., Shephard, N. (2005). How accurate is the asymptotic approximation to the distribution of realized volatility?. *Identification and inference for econometric models. A Festschrift in honour of TJ Rothenberg*, 4, 306–311.

[29] Barndorff-Nielsen, O.E., Shephard, N. (2007). Variation, jumps and high frequency data in financial econometrics. *Econometric Society Monograph*, Cambridge University Press.

[30] Bauer, R., Cosemans, M., Schotman, P. C. (2010). Conditional asset pricing and stock market anomalies in Europe. *European Financial Management*, 16(2), 165–190.

[31] Bauwens, L., and Hautsch, N., (2009), Modelling Financial High Frequency Data Using Point Processes. *Journal of Applied Econometrics*, 4, 67–93.

[32] Black, F., Jensen, M. C., and Scholes, M. (1972). The capital asset pricing model: Some empirical tests. *Studies in the theory of capital markets*, 81(3), 79–121.

[33] Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1), 116–131.

[34] Bollerslev, T., Mikkelsen, H.O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics 73*, 151–184.

[35] Bollerslev, T., Osterrieder, D., Sizova, N., Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108, 409–424.

[36] Beach, S. L. (2011). Semi variance decomposition of country-level returns. *International Review of Economics and Finance*, 20(4), 607–623.

[37] Brown, R.L., J. Durbin and J.M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of Royal Statistical Society*, 7(B), 149–192.

[38] Brownlees, C. T., and Gallo, G. M., (2006). Financial econometrics at ultra-high frequency: data handling concerns. *Computational Statistics and Data Analysis*, 51, 2232–2245

[39] Busch, T., Christensen, B.J., Nielsen, M.Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160, 48–57.

[40] Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance*, 52, 57–85.

[41] Chen, N. F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 383–403.

[42] Christensen, B.J., Nielsen, M.Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133, 343–371.

[43] Christensen, B.J., Nielsen, M.Ø. (2007). The effect of long memory in volatility on stock market fluctuations. *Review of Economics and Statistics*, 89, 684–700.

[44] Christensen, B.J., Nielsen, M.Ø., Zhu, J. (2010). Long memory in stock market volatility and the volatility-in-mean effect: The FIEGARCH-M Model. *Journal of Empirical Finance 17*, 460–470.

[45] Clements, A. E., Liao, Y. (2013). Modelling and forecasting realized volatility: getting the most out of the jump component (No. 93). National Centre for Econometric Research.

[46] Choi, K., Zivot, E., (2007). Long memory and structural breaks in the forward discount: an empirical investigation. *Journal International Money Finance.* 26, 342-?363.

[47] Corsi, F. (2009). A simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*,3, 1–23.

[48] Corsi, F., Reno, R. (2010). HAR volatility modelling with heterogeneous leverage and jumps. *Available at SSRN 1316953*, 690.

[49] Corsi, F. and Reno R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modelling. *Journal of Business and Economic Statistics*, 30(3), 368–380.

[50] Creal, D. D., Koopman, S. J., Lucas, A. (2008). A general framework for observation driven time-varying parameter models. Discussion paper.

[51] Davidson, J., Sibbertsen, P. (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of Econometrics*, 128, 253–282.

[52] Diebold, F. X. (1986). Testing for Serial Correlation in the Presence of ARCH. *In Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323, 328.

[53] Diebold, F.X., Mariano R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.

[54] Diebold, F.X., Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105, 131–159.

[55] Ding, Z., Granger C.W.J., Engle R.F. (1993). Long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83-106.

[56] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

[57] Engle, R. F., Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1), 1-50.

[58] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3), 339–350.

[59] Fama, MacBeth E. F. J. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, 71, 607- 636.

[60] Fama, E. F. and French K. R. (1992). The cross section of expected stock returns. *Journal of Finance*, 47, 427–465.

[61] Fama, E. F. and French K. R. (2016). Dissecting anomalies with a five factor model. *Review of Financial Studies*, 29, 69–103.

[62] Ferson, W. E., Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, 99(2), 385–415.

[63] Ferson, W. E., Kandel, S., and Stambaugh, R. F. (1987). Tests of Asset Pricing with Time Varying Expected Risk Premiums and Market Betas. *Journal of Finance*, 42(2), 201–220.

[64] Fox, R., Taqqu, M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian processes. *Annals of Statistics*, 14, 517–532.

[65] Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time Varying Risk Premium in Large Cross-sectional Equity Data Sets. *Econometrica*, 84(3), 985–1046.

[66] Geweke, J., Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4, 221–238.

[67] Ghysels, E. (1998). On Stable Factor Structures in the Pricing of Risk: Do Time arying Betas Help or Hurt?. *Journal of Finance*, 53(2), 549-573.

[68] Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545-1578.

[69] Giraitis, L., Kapetanios, G., and Yates, T. (2014). Inference on stochastic time varying coefficient models. *Journal of Econometrics*, 179, 46–65.

[70] Giraitis, L., Kapetanios, G., and Yates, T. (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis*, 39(2), 129-149.

[71] Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.

[72] Granger, C.W.J., Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73, 61–78.

[73] Granger, C.W.J., Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns. *Journal of Empirical Finance*, 11, 399–421.

[74] Granger, C.W.J., Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15–39.

[75] Granger, C.W.J., Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters 62*, 4, 161–165.

[76] Greene, M., Fielitz, B. (1977). Long term dependence in common stock returns. *Journal Financial Economics*, 4, 339–349.

[77] Hansen, P. R., Lunde, A. (2006b). Realized variance and market microstructure noise. *Journal of Business and economic Statistics*, 24, 127–218.

[78] Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized beta garch: A multivariate garch model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5), 774–799.

[79] Harvey, C.R. (1989). Time varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics*. 24, 289–317.

[80] Harvey, C.R. (1991). The world price of covariance risk. *Journal of Finance*, 46, 111–157.

[81] Hawkes, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58, 83–90.

[82] Hillebrand, E. (2005). Neglecting parameter changes in GARCH models, *Journal of Econometrics*, 129, 121–138.

[83] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68, 165–176.

[84] Hosoya, Y. (1997). A limit theory for long range dependence and statistical inference on related models. *Annals of Statistics*, 25, 105–137.

[85] Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770–799.

[86] Jagannathan, R., Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51(1), 3-53.

[87] Kan, R., Robotti, C., and Shanken, J. (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *Journal of Finance*, 68(6), 2617–2649.

[88] Kokoszka P, Leipus R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, 39, 231–247.

[89] Lamoureux, C. and Lastrapes W. (1990). Persistence in variance, structural GARCH model, *Journal of Business and Economic Statistics*, 8, 225–234.

[90] Lettau, M., Ludvigson, S. (2001a). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3), 815–849.

[91] Lettau, M., Ludvigson, S. (2001b). Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109(6), 1238–1287.

[92] Lewellen, J., Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2), 289–314.

[93] Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4), 587–615.

[94] Liu, L.Y., Patton, A.J., Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187, 293–311.

[95] Louzis, D.P., A.T. Vouldis (2012). A Financial Systemic Stress Index for Greece. *ECB Working Paper*, 13.

[96] Mikosch, T., Starica, C. (2004). Non stationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1), 378–390.

[97] McAleer, M., Medeiros, M.C. (2008). A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. *Journal of Econometrics*, 147, 104–119.

[98] Meddahi, N., Mykland, P., Shephard, N. (2011). Special issue on realised volatility. *Journal of Econometrics*, 160, 1.

[99] Nielsen, M.Ø., Frederiksen, P.H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24, 405–443.

[100] Ohanissian, A., Rusell, J.R. and Tsay, R.S. (2008). True or spurious long memory? A new test. *Journal of Business and economic Statistics*, 26, 161–175.

[101] Patton, A.J., Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97, 683–697.

[102] Petkova, R., Zhang, L. (2005). Is value riskier than growth?. *Journal of Financial Economics*, 78(1), 187–202.

[103] Pesaran, M. H., A. Pick, (2010). Forecasting Combination across Estimation Windows, *CESifo Working Paper Series*, 2, 22–93.

[104] Pesaran M. H. and Timmermann A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series?. *International Journal of forecasting*, 20, 411–425.

[105] Ploberger, W., Kramer, W. (1992). The *CUSUM* test with *OLS* residuals. *Econometrica*, 64, 271–285.

[106] Poskitt, D. (2007). Autoregressive approximation in nonstandard situations. The non-invertible and fractionally integrated case. *Annals of the Institute of Statistical Mathematics*, 59, 697–725.

[107] Poskitt, D. (2008). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis*, 29, 224–250.

[108] Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 4(2), 129–176.

[109] Sakoulis, G., Zivot, E., Choi, K., (2010). Time variation and structural change in the forward discount: implications for the forward rate unbiasedness hypothesis. *Journal Empirical Finance* 17, 957-966.

[110] Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

[111] Shanken, J. (1985). Multi factor CAPM or Equilibrium?: A Reply. *Journal of Finance*, 40(4), 1189-1196.

[112] Shanken, J. (1992). The current state of the arbitrage pricing theory. *Journal of Finance*, 47(4), 1569-1574.

[113] Shanken, J., Zhou, G. (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, 84(1), 40-86.

[114] Sharpe, W. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19, 425–442.

[115] Shimotsu, K., Phillips, P.C. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 33, 1890–1933.

[116] Sibbertsen, P. (2004). Long memory versus structural breaks: an overview. *Statistical Papers*, 45, 465–515.

[117] Wenger, K., Leschinski, C., Sibbertsen, P. (2018). A simple test on structural change in long memory time series. *Economics Letters*, 163, 90–94.

[118] Taylor, Steven J. (1986). Modelling financial time series. *Wiley*, New York, NY.

[119] Taylor, S. J. (1994). Modelling stochastic volatility: A review and comparative study. *Mathematical finance*, 4(2), 183–204.

[120] Yang, K., Chen, L., and Tian, F. (2015). Realized volatility forecast of stock index under structural breaks. Journal of Forecasting, 34(1), 57–82.

[121] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12, 1019–1043.

[122] Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005). A tale of two time series: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100, 1394–1411.

[123] Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2011). Edgeworth expansions for realized volatility and related estimators. *Journal of Econometrics*, 160(1), 190–203.

[124] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics*, 14, 45–52.

[125] Zolotoy, L. (2011). Earnings News and Market Risk: Is the Magnitude of the Post earnings Announcement Drift Underestimated?. *Journal of Financial Research*, 34(3), 523–535.